Technical Report 792

1

Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1986 Fiscal Year

John P. Campbell, Editor
Human Resources Research Organization

Selection and Classification Technical Area

Manpower and Personnel Research Laboratory

U. S. Army

Research Institute for the Behavioral and Social Sciences

May 1988

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) HumRRO IR-PRD-87-10	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Technical Report 792
6a. NAME OF PERFORMING ORGANIZATION Human Resources Research Organization 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences
6c. ADDRESS (City, State, and ZIP Code) 1100 South Washington Street Alexandria, VA 22314-4499	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600
8a. NAME OF FUNDING/SPONSORING ORGANIZATION -- 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA 903-82-C-0531
8c. ADDRESS (City, State, and ZIP Code) --	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 6.37.31.A PROJECT NO. 2Q263731A792 TASK NO. 2.3.2 WORK UNIT ACCESSION NO. C.1

11. TITLE (Include Security Classification)
Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1986 Fiscal Year

12. PERSONAL AUTHOR(S)
John P. Campbell, Editor (HumRRO)

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 10/85 TO 09/86	1988, May	230

16. SUPPLEMENTARY NOTATION Prepared under Project A: Improving the Selection, Classification, and Utilization of Army Enlisted Personnel (Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, U.S. Army Research Institute).

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Army-wide measures; Criterion measures; First-tour evaluation; Personnel classification; Personnel selection, (Continued)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report describes the research performed during the 4th year (FY86) of Project A, the Army's long-term program to develop a complete system for selecting and classifying all entry-level enlisted personnel. During the 4th year, the wide variety of predictor and criterion measures that had been developed and field tested during the first 3 years were administered to 9,500 soldiers in the Concurrent Validation phase of the project. This report describes the data collection and analysis for this phase, which is being followed by a Longitudinal Validation phase in which approximately 50,000 soldiers in 21 Military Occupational Specialties (MOS) will be tested on entry into the Army and during subsequent first-tour job performance. This report is supplemented by an ARI Research Note (in preparation), which contains a number of technical papers prepared during the year on specialized aspects of the project.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Lawrence M. Hanser	22b. TELEPHONE (Include Area Code) (202) 274-8275 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON Technical Director	WM. DARRYL HENDERSON COL, IN Commanding

Research accomplished under contract
for the Department of the Army

Human Resources Research Organization

Technical review by

Darlene M. Olson
Michael G. Rumsey
Deirdre Knapp
Leonard A. White
Clint B. Walker

Dist

A-1

QUALITY INSPECTED 2

NOTICES

~~DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POT, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600.~~

FINAL DISPOSITION: *This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.*

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

ARI Technical Report 792

18. Subject Terms (Continued)

Predictor measures
Project A
Rating scales
Soldier effectiveness

Technical Report 792

Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1986 Fiscal Year

John P. Campbell, Editor
Human Resources Research Organization

Selection and Classification Technical Area
Lawrence M. Hanser, Chief

Manpower and Personnel Research Laboratory
Newell K. Eaton, Director

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel

Department of the Army

May 1988

Army Project Number 2Q263731A792

Manpower and Personnel

FOREWORD

This document is a description of the research effort during the 4th year (Fiscal Year 1986) of the Army's current, large-scale effort for improving the selection, classification, and utilization of Army enlisted personnel. The thrust for the project came from the practical, professional, and legal need to validate the Armed Services Vocational Aptitude Battery (ASVAB--the current U.S. military selection/classification test battery) and other selection variables as predictors of training and performance. The present report describes the initial experience in determining how well the new selection/classification tests predict job performance.

The portion of the effort described herein is devoted to the development and validation of Army Selection and Classification Measures, referred to as "Project A," which is being conducted under contract to the Selection and Classification Technical Area (SCTA) of the Manpower and Personnel Research Laboratory (MPRL) at the U.S. Army Research Institute for the Behavioral and Social Sciences. This research supports the MPRL and SCTA mission to improve the Army's capability to select and classify its applicants for enlistment or reenlistment by ensuring that fair and valid measures are developed for evaluating applicant potential based on expected job performance and utility to the Army.

Project A was authorized through a Letter, DCSOPS, "Army Research Project to Validate the Predictive Value of the Armed Services Vocational Aptitude Battery," effective 19 November 1980; and a Memorandum, Assistant Secretary of Defense (MRA&L), "Enlistment Standards," effective 11 September 1980.

In order to ensure that Project A research achieves its full scientific potential and will be maximally useful to the Army, a governance advisory group comprised of Army general officers, interservice scientists, and experts in personnel measurement, selection, and classification was established. Members of the last component provide guidance on technical aspects of the research, while general officer and interservice components oversee the entire research effort; provide military judgment; provide periodic reviews of research progress, results, and plans; and coordinate within their commands. Members of the General Officers' Advisory Group during the period covered by this report included MG W. G. O'Leksy (DMPM) (Chair), MG C. F. Briggs (FORSCOM, DCSPER), BG W. C. Knudson (DCSOPS), BG F. M. Franks, Jr. (USAREUR, ADCSOPS), and MG C. H. Corns (TRADOC, DCS-T). The General Officers' Advisory Group was briefed in November 1985 on the status of the concurrent validation data collection and the preliminary results. Members of Project A's Scientific Advisory Group (SAG) guide the technical quality of the re- search. During the period covered by this report, they included Drs. Philip Bobko, Thomas Cook, Milton Hakel (Chair), Lloyd Humphreys, Lawrence Johnson, Robert Linn, Mary Tenopyr, and Jay Uhlaner. The SAG was briefed in March 1986 on the rationale and rules adopted for preparing the concurrent validation criterion data for analysis and in September 1986 on the concurrent validation analysis results.

A comprehensive set of new selection/classification tests and job performance/training criteria have been developed and field tested, and the revised tests have been administered in a large-scale concurrent validation data collection effort. Results will be used to link enlistment standards to required job performance standards and to more accurately assign soldiers to Army jobs.

EDGAR M. JOHNSON
Technical Director

PREFACE

The Army Selection and Classification Project (Project A) has completed the 4th year (Fiscal Year 1986) of its 9-year schedule of research. Most of the project's first 3 years were taken up with planning, literature reviews, extensive job analyses, and careful development of both a comprehensive battery of new selection/classification tests and a comprehensive array of job performance criterion measures. Both development efforts were as thorough and state-of-the-art as we could make them.

The 4th year of the project was the first opportunity to estimate how well the Armed Services Vocational Aptitude Battery (ASVAB) and the new Project A selection/classification tests predict job performance. The validity data were provided by the assessment of 9,500 job incumbents in a concurrent validity design.

The Annual Report, 1985 Fiscal Year was designed as a complete account of the first 3 years of the project and documented the development of the selection/classification tests and the job and training performance measures. The present report is an account of the Concurrent Validation data collection and the basic validation analysis. Taken together, these two Annual Reports provide a complete account of the entire project through its 4th year. They are backed up by even more detailed field test reports, field test report appendixes, data base documentation, literature reviews, and other technical reports.

The procedure that was followed to create this Annual Report, 1986 Fiscal Year is similar to that used for the FY85 Annual Report. That is, various papers originally written for presentation at review meetings, conferences, or other special purposes were edited and revised to conform to the needs of the Annual Report. Gaps in the story were filled in as needed. Credit for the original material is given on the first page of each chapter; the sources upon which the chapter is based are described and the authors of the original documents are listed.

At this point all of us on the project are still awestruck by its proportions and we are absolutely amazed that it has met its objectives, produced the products it said it would produce, and is now beginning to produce the components of a new selection/classification system. We hope the FY86 Annual Report provides interesting reading for all concerned.

John P. Campbell
Editor

IMPROVING THE SELECTION, CLASSIFICATION, AND UTILIZATION OF ARMY ENLISTED PERSONNEL: ANNUAL REPORT, 1986 FISCAL YEAR

EXECUTIVE SUMMARY

Requirement:

Project A is a comprehensive, long-range U.S. Army program to develop an improved personnel selection and classification system for enlisted personnel. The system encompasses 675,000 persons and several hundred Military Occupational Specialties (MOS). The objectives are (a) to validate existing selection measures against both existing and project-developed criteria and to develop new measures; and (b) to validate early criteria (e.g., performance in training) as predictors of later criteria (e.g., job performance ratings), in order to improve reassignment and promotion decisions.

Procedure:

With the Deputy Chief of Staff for Personnel (DCSPER) as sponsor, work on the 9-year project was begun in 1982. In the first stage, file data from FY81/82 Army accessions were used to explore the relationships between the scores applicants made on the Armed Services Vocational Aptitude Battery (ASVAB) and their later performance in training and first-tour skill tests. The second stage was executed with FY83/84 accessions in 19 MOS selected as representative of the Army's 250+ entry-level MOS, and accounting for 45 percent of Army accessions. A preliminary battery of perceptual, spatial, temperament, interest, and biodata predictor measures was tested with several thousand soldiers as they entered four MOS; later versions were pilot tested and field tested with nine MOS. The resulting predictor battery and a comprehensive set of job knowledge tests, hands-on job samples, and performance ratings were administered to 9,500 soldiers in 19 MOS in the "Concurrent Validation." In the third stage, the measures, refined with experience, will be used to test about 50,000 soldiers across 21 MOS in the FY86/87 predictor battery administration and subsequent measurement of first-tour performance. About 3,500 soldiers are expected to be available for second-tour performance measurement in FY91.

Findings:

Analysis of results from the Concurrent Validation testing led to improved understanding of the major factors that contribute to good performance and how to measure them. The most effective of the predictor and criterion measures developed and tested during the first 4 years of Project A are now being used in the "Longitudinal Validation" phase, which began with recruits entering the Army during FY86 and will continue with measurement of first-tour performance in subsequent years.

Utilization of Findings:

The full array of Project A selection/classification measures of training and job performance is being used in both current and long-range research programs expected to make the Army more effective in matching first-tour enlisted manpower requirements with available personnel resources.

IMPROVING THE SELECTION, CLASSIFICATION, AND UTILIZATION OF ARMY ENLISTED PERSONNEL: ANNUAL REPORT, 1986 FISCAL YEAR

OVERVIEW OF PROJECT A

Project A is a comprehensive long-range research and development program the U.S. Army has undertaken to develop an improved personnel selection and classification system for enlisted personnel. The Army's goal is to increase its effectiveness in matching first-tour enlisted manpower requirements with available personnel resources through use of new and improved selection/classification tests that will validly predict carefully developed measures of job performance. The project addresses the Army's 675,000-person enlisted personnel system encompassing several hundred Military Occupational Specialties (MOS).

The program began in 1980, when the U.S. Army Research Institute (ARI) started planning the extensive research needed to develop the desired system. In 1982 ARI selected a consortium, led by Human Resources Research Organization (HumRRO) and including American Institutes for Research (AIR) and Personnel Decisions Research Institute (PDRI), to undertake the 9-year project. It is utilizing the services of 40 to 50 ARI and consortium researchers working collegially in a variety of professional specialties. The Project A objectives are as follows:

- Validate existing selection measures against existing and project-developed criteria (including both Army-wide job performance measures based on rating scales and direct hands-on measures of MOS-specific task performance).

- Develop and validate new selection and classification measures.

- Validate intermediate criteria, such as training performance, as predictors of later criteria, such as job performance, so that better informed decisions on reassignment and promotion can be made throughout a soldier's career.

- Determine the relative utility to the Army of different performance levels across MOS.

- Estimate the relative effectiveness of alternative selection and classification procedures in terms of their validity and utility for making decisions.

The research design incorporates three main stages of data collection and analysis in an iterative progression of development, testing, evaluation, and further development of selection/classification instruments (predictors) and measures of job performance (criteria). In the first iteration, file data from fiscal years (FY) 1981/1982 were evaluated to explore relationships between scores of applicants on the Armed Services Vocational Aptitude Battery (ASVAB), and their later performance in training and their scores on first-tour Skill Qualification Tests (SQT).

For the ensuing research, 19 Military Occupational Specialties were selected as a representative sample of the Army's 250+ entry-level MOS. The selection was based on an initial clustering of MOS derived from rated similarities of job content. These MOS account for about 45 percent of Army accessions and provide sample sizes large enough so that race and sex fairness can be empirically evaluated in most MOS.

In the second iteration, a Concurrent Validation design was executed with FY83/84 accessions. A Preliminary Battery of perceptual, spatial, temperament, interest, and biodata predictor measures was developed and tested with several thousand soldiers as they entered four MOS. The data from this sample were then used to refine the measures, with further exploration of content and format. The revised set of measures was field tested to assess reliabilities, "fakability," practice effects, and other factors. The resulting predictor battery, or Trial Battery, was administered together with a comprehensive set of job performance indexes based on job knowledge tests, hands-on job samples, and performance rating measures, in the Concurrent Validation during the summer and fall of 1985.

On the basis of testing experience, the Trial Battery was revised as the Experimental Predictor Battery, which in turn is being administered in the Longitudinal Validation stage (third iteration) beginning in the late summer of 1986. Three MOS have been added to the original 19 (19K, 29E, and 96B), and one of the original MOS was dropped (76W). All measures are being administered in a true predictive validity design. About 50,000 soldiers across 21 MOS will be included in the FY86-87 administration and subsequent first-tour measurement. About 3,500 of these soldiers are expected to be available for second-tour performance measurement in FY91.

For administrative purposes, Project A is divided into five research tasks: Task 1, Validity Analyses and Data Base Management; Task 2, Developing Predictors of Job Performance; Task 3, Developing Measures of School/Training Success; Task 4, Development Measures of Army-Wide Performance; and Task 5, Developing MOS-Specific Performance Measures.

Activities during the first 3 years of Project A were reported as follows: FY83, ARI Research Report 1347 and its Technical Appendix, ARI Research Note 33-37; FY84, ARI Research Report 1393 and two related reports, ARI Technical Report 660 and ARI Research Note 85-14; FY85, ARI Technical Report 746 and ARI Research Note 87-54. The present FY86 report is supplemented by ARI Research Note (in preparation). These reports list other publications on specific Project A activities.

IMPROVING THE SELECTION, CLASSIFICATION, AND UTILIZATION OF ARMY ENLISTED PERSONNEL: ANNUAL REPORT, 1986 FISCAL YEAR

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

THE HIGHLIGHTS OF FISCAL YEAR 1986 AND THE OBJECTIVES OF THIS REPORT[1]

This report is intended to be a summary of the major activities in the Project A research program during fiscal year 1986 (FY86). Prior year Annual Reports for FY83, FY84, and FY85 respectively concentrated on a description of the research planning and basic preparation (FY83), the initial stages of the development of new predictor and criterion tests (FY84), and a comprehensive report of the full predictor/criterion development and field testing (FY85).

This report for FY86 focuses on the Concurrent Validation (CV) portion of Project A and the basic analyses of CV data while, at the same time, the planning, preparation, and coordination were being accomplished for the equally comprehensive Longitudinal Validation (LV) data collections scheduled for FY87, FY88, and FY89. The Project A research design is summarized in Figure 1.1. The content of this report is taken from the activities within the dotted lines. The reader is referred to earlier reports for a complete description of the project design.

Briefly restated, the operational objectives of Project A are to:

- Develop new measures of job performance that the Army can use as criteria against which to validate selection/classification measures.
- Validate existing selection measures against both existing and project-developed criteria.
- Develop and validate new selection and classification measures.
- Develop a utility scale for different performance levels across military occupational specialties (MOS).
- Estimate the relative effectiveness of alternative selection and classification procedures in terms of their validity and utility.

In addition, a number of related and derivative research objectives have been addressed in the overall research program. Salient data and analyses in regard to those objectives are also summarized in this report.

---

[1]The major sections of Chapter 1 were originally drafted by Dr. Marvin Goer, Project Director Emeritus, who retired from HumRRO and Project A on 30 January 1987.

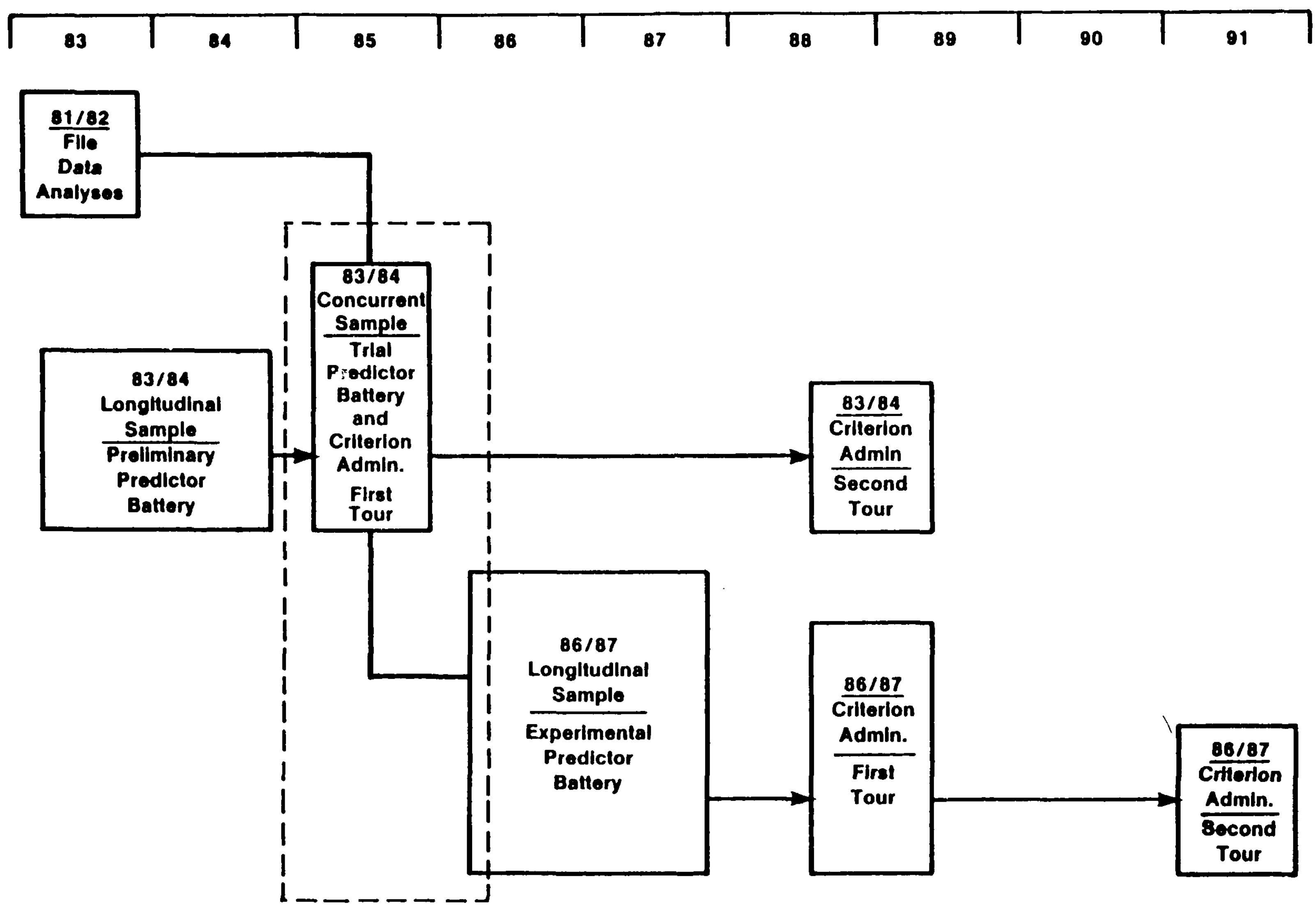


Figure 1.1. The overall research design for Project A.

STATE OF PROJECT A AS OF 30 SEPTEMBER 1985

As described in the Annual Report for FY85, the state of Project A at the beginning of FY86 was as follows:

1. All development work for the new predictor measures that comprised the Trial Battery had been completed; that is, a 4-hour battery of new selection tests and inventories had been carefully developed and fully field tested. A complete report of the Trial Battery development is given in Peterson (1987), and a shorter version can be found in Campbell (1987). The predictor array is listed in Chapter 2, which describes the Concurrent Validation design.

2. All development work on the complete array of training and job performance measures had also been completed. The 12-hour assessment package includes a 4-hour hands-on (job sample) test, 4 hours of knowledge tests, multiple rating scales, questionnaires, and self-reports of personnel records information. A complete description of the three years of criterion development and field test work is given in a series of reports (Campbell, Campbell, Rumsey, & Edwards, 1986; Davis, Davis, Joyner, & de Vera, 1987; Pulakos & Borman, 1986a; Toquam et al., 1985; and Riegelhaupt, Harris, & Sadacca, 1987). An abridged description is given in Campbell (1987). The full criterion array used in the Concurrent Validation is listed in the next chapter.

3. The data collection procedure had been designed, the data collection teams had been assembled and trained, and approximately 65 percent of the Concurrent Validation data had been collected. The data collection design and procedure are summarized in the next chapter.

REVIEW OF FY86 SIGNIFICANT EVENTS

FY86 was an action-packed year. Some of the more significant highlights are as follows:

- Dr. Lawrence M. Hanser was officially selected to head the Army Research Institute for the Behavioral and Social Sciences (ARI) Selection and Classification Technical Area and was designated as the Contracting Officer's Technical Representative (COTR) for Project A. He succeeds Dr. N. K. Eaton, who was officially selected to be Director of the ARI Manpower and Personnel Research Laboratory (MPRL).

  Dr. Eaton also assumed the chairmanship of the Interservice Advisory Group (ISAG) for Project A. Dr. Lawrence Johnson was invited by ARI to join the Project A Scientific Advisory Group (SAG), further strengthening the expertise available to the project in addressing broad policy implications entailed in implementing potential improvements in selection and classification of Army enlisted personnel.

The composition of the Project A governance groups and the organization of the research and oversight staff at the end of FY86 are shown in Figures 1.2 and 1.3.

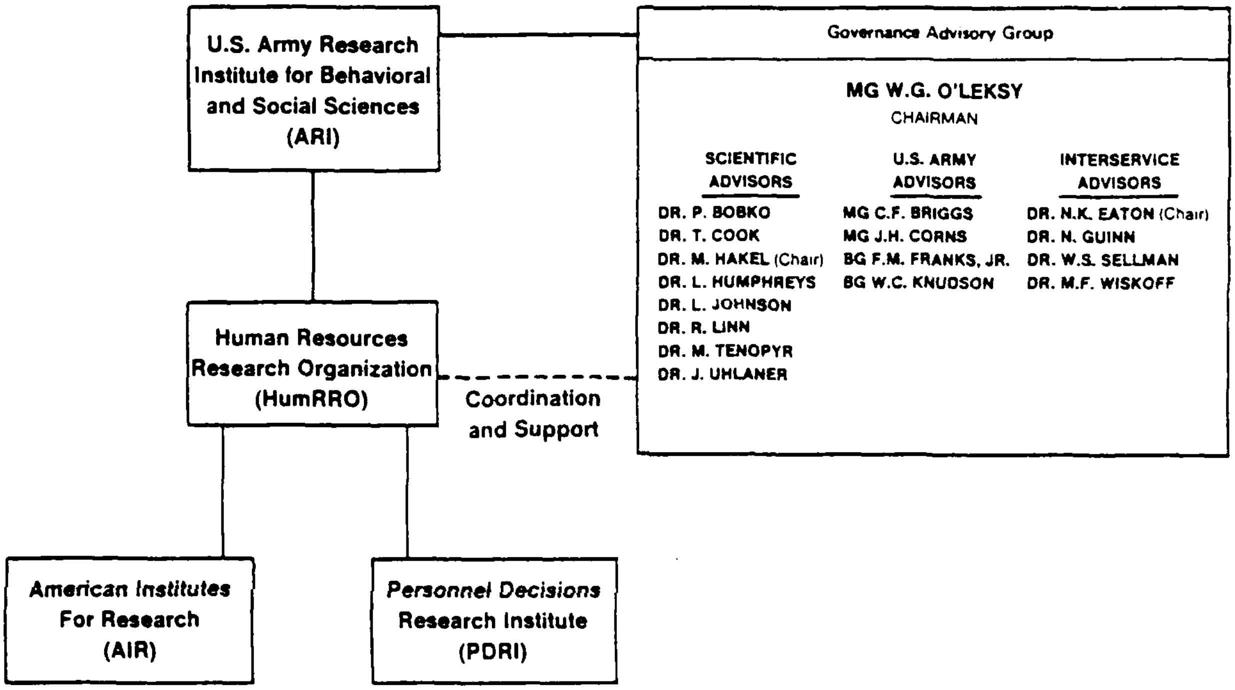


Figure 1.2. Project A organization as of 30 September 1986.

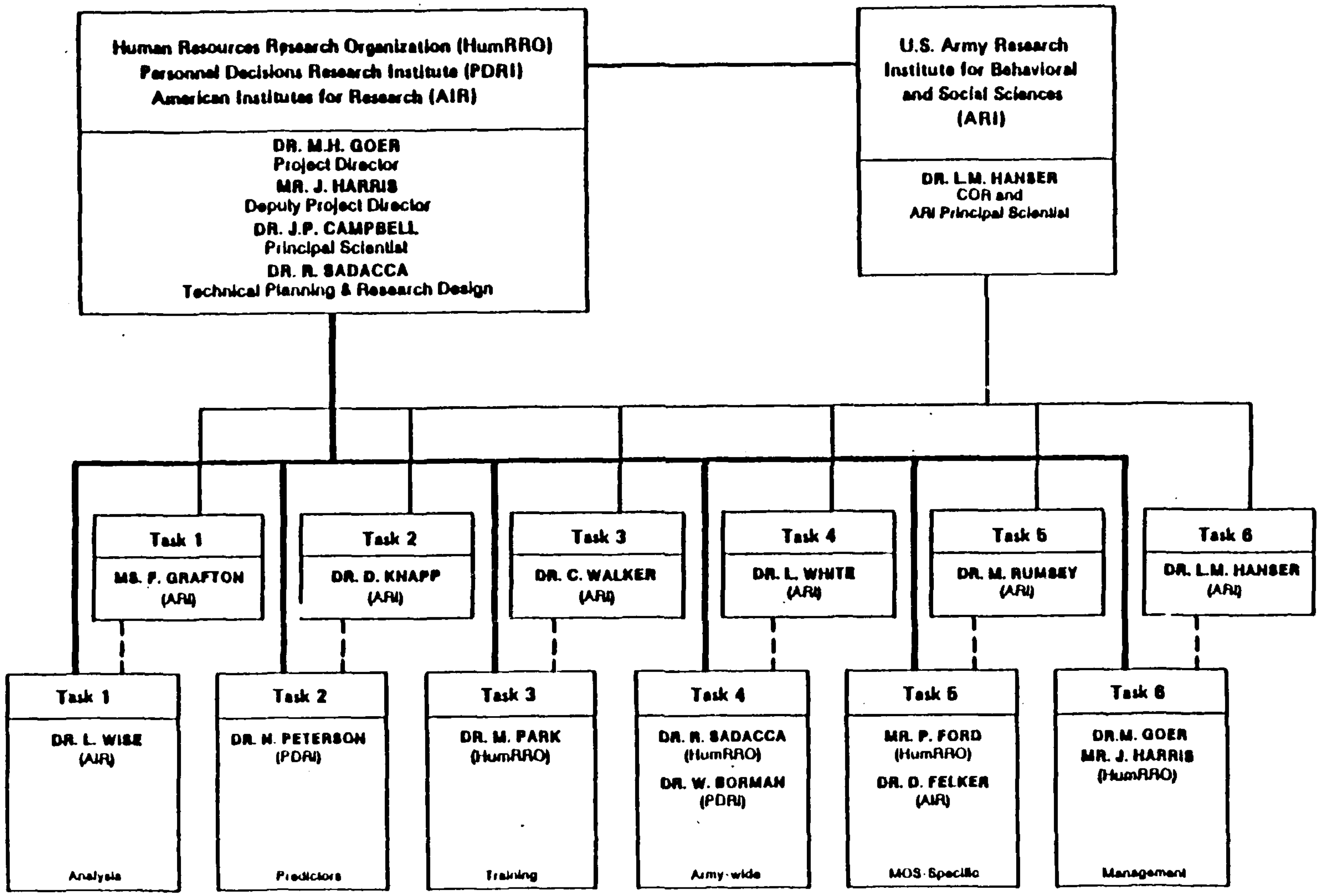


Figure 1.3. Project A Management Group as of 30 September 1986.

o FY86 began as the Concurrent Validation data collection was nearing completion (work was finished at the last two sites--Fort Ord and Fort Knox--late in November 1986). Despite strong pressure for the release of partial or incompletely analyzed results, ARI, with the concurrence of the SAG, determined that CV results would not be released or reported until the project scientists had thoroughly examined and analyzed the data. Accordingly, rigorous rules and procedures were developed for dealing with the "missing data," and for forming criterion and predictor constructs, and a comprehensive analysis program was developed, reviewed and approved, and subsequently executed. The body of this report extensively describes that process and its outcome.

o Important decisions were taken in respect to the scope and focus of the planned Longitudinal Validation research program. Two military occupations (MOS) were added to the job sample--Electronics Repair (29E) and Intelligence Analyst (96B)--to cover job families that had not been adequately represented in the CV sample of jobs, while one job--Petroleum Supply Specialist (76W)--was dropped because that job family was deemed to be well represented by the remaining MOS. The resultant number of military occupations planned for the LV data collections therefore expanded to 21.

  Another decision was to administer the predictor measures as early in a soldier's first tour as was feasible. Arrangements were coordinated, through the U.S. Army Training and Doctrine Command (TRADOC), to administer the predictor battery to soldiers while they were in process in the reception stations, and the school knowledge tests to soldiers at the end of Advanced Individual Training (AIT) or One Station Unit Training (OSUT). It was also decided that contractor test administrators would be used and that all occupations would be tested for a full year, yielding a total LV sample estimated at more than 50,000 cases. To accomplish the testing at reasonable throughput rates, 34 additional psychomotor/ perceptual testing devices were authorized and acquired.

o Development work on second-tour criterion measures was deferred to FY87, pending development of information on the salient dimensions of performance and the number of occupations for which reasonable samples could be obtained. Based on those data, informed decisions then could be made on the scope of criterion measure development.

o An increased level of interaction and coordination of research and development was undertaken between Project A and a separate ARI priority project, Project B - Development of a Computerized Personnel Allocation System. It has been ARI's intention to have Project A data and results used to inform the development of computerized systems for Army personnel managers in recruiting, incentive systems, classification and assignment, and soldier retention programs. Project B is engaged in developing a set (system) of computer models that permit alternative personnel policy options and outcomes to be evaluated, thus providing information for the development and adoption of Army personnel policies.

MEETINGS, REVIEWS, AND BRIEFINGS

The Project A staff continued their longstanding program of meticulously coordinating, reviewing, and subsequently communicating research in progress, as well as results that could be confidently released to the Army, other government agencies, and the scientific community in general. This program was accomplished through the mechanism of regularly scheduled meetings with the SAG and frequent In-Progress Reviews (IPRs), and through participation in national professional conferences. The following list summarizes the FY86 program.

<u>October</u>

Projects A/B Coordination Meeting - Status of both projects was reviewed. Convergence of schedules and interface between the projects were evaluated.

Scientific Advisory Group (SAG) - Status of CV data collection was reviewed. Policy of completing "clean-up" and analyses before evaluating CV results was endorsed. The Group recommended the addition of two MOS while deleting one MOS.

<u>November</u>

General Officers Advisory Group (GAG) - The GAG endorsed changes in the occupational sample recommended by SAG. They suggested that TRADOC designate the proponent for Army-wide measures, and that LV predictor testing be performed at the reception stations.

<u>December</u>

Analysis Group (Task 1) IPR - CV data analyses plans were reviewed. Plans for dealing with missing data and incomplete tests were presented and reviewed. An initial model for the latent structure for criterion variables was proposed and reviewed.

<u>January</u>

Management Group (Task 6) IPR - FY86 work plans, budgets, and priorities were discussed and determined. The need to update the Research Plan was presented and discussed.

<u>February</u>

Criterion Group (Tasks 3, 4, 5) IPR - Specific rules were reviewed for analyzing outliers, dealing with missing CV criterion data, forming criterion scores, formulating criterion constructs, and weighting the criterion factors.

March

Predictor Group (Task 2) IPR - Scoring methods for cognitive, non-cognitive, and computerized measures were reviewed. Improvements in test instructions were suggested. LV predictor data collection plans were reviewed and discussed.

SAG Meeting - Rules and procedures for editing and preparing the CV criterion data for analysis were reviewed, discussed, and approved.

May

Analysis Group (Task 1) IPR - Progress in reducing criterion constructs to a manageable set was discussed. Validity estimation alternatives were examined. Various methods for effectively communicating the effects of increased validities were reviewed.

July

Management Group (Task 6) IPR - Status of project resources was reviewed. Forecasts of resources required to complete the existing Research Plan were discussed. Information on options was requested. Requirement to develop briefing for CINC/USAREUR was discussed.

August

CINC/USAREUR Briefing - Following a preview by Mr. Spurlock, Assistant Secretary of the Army, LTG Elton, and MG O'Leksy, Dr. Hanser and Dr. Eaton provided a summary briefing on Project A status and results to date, including results of the CV data analysis and interpretation, for GEN Otis, CINC/USAREUR. GEN Otis indicated his satisfaction and expressed his continued support.

September

SAG/Criterion Group (Tasks 3, 4, 5) IPR - In addition to a review of the results of the CV analyses, and a presentation on Project B, a number of salient research issues confronting Project A were discussed. These included issues of sample sizes for second-tour measurement, the content and method of second-tour measurement, continuation of psychomotor/perceptual data collection, AVOICE implementation, progress in construct weighting, and the interface/integration of Projects A/B. Dr. Eaton observed that Project A had successfully met one of its major objectives in the validation of ASVAB against job performance and that ARI was pleased with both the thoroughness of the research and the credibility of the data. The SAG deferred specific recommendations on second-tour issues pending a review of data obtained in analysis of job activities for the Project A 21-job sample.

It was suggested that "second tour" be operationally defined as soldiers who have been in the Army between 13 and 20 quarters and who are performing senior Skill Level 1, 2, or 3 duties. Additional information will be gathered during the job analyses to help in this decision.

SAG noted a need to assure closer coordination and integration of Project A/B research and recommended that ARI establish procedures to insure attainment of this objective.

MAJOR PROJECT ACTIVITIES FOR FY87

Work will continue in accordance with the current Research Plan and schedule. As has been noted many times, Project A is cohort driven and we cannot stop to catch our breath or rest on laurels. During FY87 we must complete the data collection on the FY86/87 sample of recruits (N = 50,000), using the Experimental Predictor Battery. Additional analyses and reports deriving from the CV data set also will continue to be produced.

Two of the original tasks--Developing Predictors of Job Performance (Task 2) and Developing Measures of School/Training Success (Task 3)--will phase out and those staffs will be engaged in producing the final reports on those activities. The LV data collection (predictor and school knowledge testing) will continue through the year. Those data will be subjected to the same scrupulous treatment that has been our standard throughout.

Intensive job analyses will be conducted for the occupations now comprising the Project A sample to inform pending decisions on the content, performance dimensions, possible measurement methods, and generalizability of criteria for "second-tour" criteria to be identified for development. Application is essentially a scientific and research design matter. The number of occupations, the scale of criterion instrument development, and the scale of the data collections will be constrained within economic and budgetary parameters.

It is anticipated that the current Research Plan will be updated and revised. Necessary planning and precise scheduling will be coordinated and approved for managing the development of measures, and conducting the LV criterion data collections for both first- and second-tour on-the-job performance, as well as the second-tour CV criterion administration.

OBJECTIVES OF THIS REPORT

The purpose of this FY86 Annual Report is to describe the Concurrent Validation portion of the long-range Project A program. The chapters that follow will summarize (a) the measures that were used; (b) the sample and data collection procedure; (c) the operations required to edit the data and prepare the data files for analysis; (d) the analysis of the predictors and the development of the predictor construct scores; (e) the analysis of the criterion data and the development of criterion scores; (f) the development and testing of a job performance model; (g) the analysis of the basic validity data; (h) the development of criterion composites via expert judgment scaling workshops; and (i) future plans. This report is supplemented by an ARI Research Note (in preparation), which contains a number of technical papers prepared during the year on specialized aspects of the project.

Chapter 2

THE CONCURRENT VALIDATION: SAMPLES AND PROCEDURES[1]

The original Project A Research Plan specified a Concurrent Validation target sample size of 600-700 job incumbents for each of 19 MOS. Compared to previous designations, the nomenclature for MOS groupings changed slightly for the Concurrent Validation phase. The previously designated Batch A and Batch B MOS are now known collectively as Batch A; they are the nine MOS that were used in the criterion field tests. The remaining 10 MOS are still designated as Batch Z. The Batch A and Batch Z MOS are listed in Table 2.1.

Table 2.1

MOS Used in the Concurrent Validation Phase of Project A

Batch A		Batch Z	
11B	Infantryman	12B	Combat Engineer
13B	Cannon Crewman	16S	MANPADS Crewman
19E	Armor Crewman	27E	TOW/Dragon Repairer
31C	Single Channel Radio Operator	51B	Carpentry/Masonry Specialist
63B	Light Wheel Vehicle Mechanic	54E	Chemical Operations Specialist
64C	Motor Transport Operator	55B	Ammunition Specialist
71L	Administrative Specialist	67N	Utility Helicopter Repairer
91A	Medical Specialist	76W	Petroleum Supply Specialist
95B	Military Police	76Y	Unit Supply Specialist
		94B	Food Service Specialist

The Research Plan called for a CV starting date of 1 May 1985, using procedures that had been tried out and refined during the predictor and criterion field tests. The plan specified 13 data collection sites in the Continental United States (CONUS) and two in Europe (USAREUR). The number of sites was the maximum that could be visited within the project's budget constraints, which dictated that sites be chosen to maximize the probability of obtaining the required sample sizes.

[1]The material in this chapter was assembled by the editor from a number of briefings and Scientific Advisory Group meetings. The description of the data collection procedure was originally drafted by James H. Harris.

The data collection actually began 10 June 1985 and was concluded 13 November 1985. The schedule, by site, is shown in Figure 2.1. Although the starting date was slightly later than planned, it was still within the permissible "window" that would maintain the project's original schedule.

Fort Lewis	**10 Jun - 18 Jul**	**Fort Knox**	**15 Aug - 13 Sep**
Fort Benning	**17 Jun - 25 Jul**	**Fort Sill**	**3 Sep - 4 Oct**
Fort Riley	**8 Jul - 9 Aug**	**Fort Campbell**	**3 Sep - 14 Oct**
Fort Carson	**8 Jul - 23 Aug**	**Fort Polk**	**30 Sep - 13 Nov**
Fort Hood	**8 Jul - 27 Aug**	**Fort Bliss**	**1 Oct - 31 Oct**
Fort Stewart	**22 Jul - 30 Aug**	**Fort Ord**	**7 Oct - 15 Nov**
Fort Bragg	**1 Aug - 13 Sep**	**USAREUR**	**12 Jul - 8 Oct**
			12 Jul - 9 Aug
			20 Sep - 18 Oct

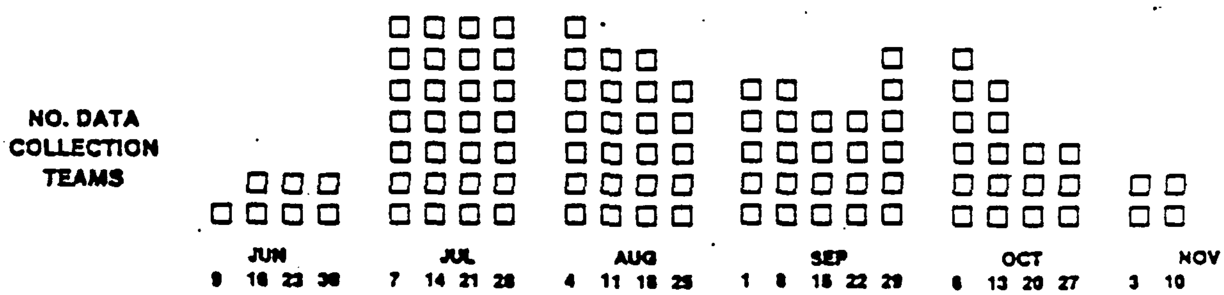


Figure 2.1. Concurrent Validation data collection.

The data were collected by on-site teams made up of project staff. Each square in Figure 2.1 represents one week of one team's time. For example, during the week of 7 July seven teams were operating, one at each of seven posts.

The basic sampling plan, team training, and data collection procedures are summarized in this chapter, and the planned data analyses are outlined. The results of the analytic steps that have been completed are summarized in following chapters.

CONCURRENT VALIDATION SAMPLE

The general sampling plan was to use the Army's World-Wide Locator System to identify all the first-term enlisted personnel in Batch A or Batch Z MOS at each specified post who entered the Army between 1 July 1983 and 30 July 1984. If possible, the individual's unit identification was also to be retained. The steps described below were then followed to the extent possible. The intent was to be as representative as possible while preserving enough cases within units to provide a "within rater" variance estimate for the supervisor and peer ratings.

Sampling Plan

Ideally, we wanted to identify the subset of MOS (within the sample of 19) for which it would be possible to actually sample people within units at specific posts. That is, given the entry date "window" and given that only 50-75 percent of the people on any list of potential subjects could actually be found and tested, what MOS are large enough to permit sampling to actually occur? Ideally, we wanted to sample 4-6 units from each post and 6-12 people from each unit. For the total concurrent sample this would provide enough units to average out or account for differential training effects and leadership climates, while still providing sufficient degrees of freedom for investigating within-group effects such as rater differences in performance appraisal.

The ideal implementation would have been to first obtain the Alpha Roster list of the total population of people at each post who were in the 19 MOS and who fit our "window". The lists would be sent to HumRRO for the following steps: (a) For each MOS, randomize units and randomize names within units. (b) Select a sample of units at random, selecting enough to allow for some units being truly unobtainable at the time of testing. (c) Instruct the Point-of-Contact (POC) at the post to obtain the required number of people by starting at the top of the list and working down (as in the Batch A field test) within each of the designated units. (d) In those MOS for which unit sampling is not possible, create a randomized list of everyone on the post who fits the "window"; instruct the POC to obtain the required number by going down the list from top to bottom. (e) If it is not possible to randomize names at the post, first use the World-Wide Locator to obtain a randomized list, carry the list to the post, and use it to sample names from units drawn from a randomized list of units. If there are only 6-8 units on the post, then no sampling of units is possible; use them all.

In practice, the ideal plan was not feasible at all the installations. The most frequent procedure was to give the sampling plan to the POC and assist him or her in obtaining the required number of people in the most representative way possible.

Actual Samples Obtained

The final sample sizes are shown by post and by MOS in Table 2.2. Note that it was not always possible in all MOS to find as many as 600 incumbents with the appropriate accession dates at the 15 sites. Some MOS simply are not that large.

Table 2.2

Concurrent Validation Sample Soldiers by MOS by Location

	Batch A MOS									Batch Z MOS											
Location	11B	13B	19E	31C	63B	64C	71L	91A	95B	12B	16S	27E	51B	54E	55B	67N	76W	76Y	94B	Total	% Total
Fort Benning	45	23	41	7	13	39	16	9	13	13	15	3	0	12	18	9	13	15	12	316	3.35
Fort Bliss	0	20	30	15	61	45	17	0	44	15	5	2	0	14	0	12	6	31	30	347	3.68
Fort Bragg	68	46	0	0	37	25	41	10	72	82	75	13	19	72	20	7	42	39	62	730	7.74
Fort Campbell	90	28	0	20	60	45	54	44	43	90	23	10	0	32	18	42	51	61	46	757	8.03
Fort Carson	60	50	77	30	49	53	30	33	46	49	57	13	0	25	7	0	23	40	47	689	7.31
Fort Hood	26	56	0	30	40	28	38	50	60	51	60	4	12	62	36	44	72	41	57	767	8.13
Fort Knox	29	32	111	16	38	48	22	45	31	43	10	6	0	8	12	0	10	29	34	524	5.56
Fort Lewis	75	46	13	11	43	46	23	27	56	27	25	1	11	51	31	20	48	41	36	631	6.69
Fort Ord	30	0	0	14	30	42	31	43	51	51	7	8	1	4	7	15	23	40	28	425	4.51
Fort Polk	73	47	19	29	47	47	18	46	44	60	45	9	8	16	7	23	26	51	35	648	6.87
Fort Riley	30	43	55	27	26	45	35	30	40	31	20	8	8	25	52	0	20	39	45	579	6.14
Fort Sill	0	108	0	20	43	51	44	0	29	42	11	0	0	0	0	15	7	35	32	437	4.63
Fort Stewart	44	46	39	17	28	51	31	45	45	30	39	9	8	17	29	26	44	34	35	617	6.54
USUREUR	132	122	120	130	122	121	114	119	118	120	78	61	41	96	54	63	105	134	113	1963	20.80
Total	702	667	503	366	637	686	514	501	692	704	470	147	108	434	291	276	490	630	612	9430	
% Total	7.44	7.07	5.3[illegible]	3.88	6.76	7.27	5.45	5.31	7.34	7.47	4.90	1.56	1.15	4.60	3.09	2.93	5.20	6.68	6.49		

PREDICTOR AND CRITERION MEASURES

The full array of predictor and criterion measures used in the Concurrent Validation is described at some length in the FY85 Annual Report and in full detail in the development and field test reports for each major type of instrument. The variables in each domain are listed in Tables 2.3 and 2.4. In the Concurrent Validation one-half day was devoted to predictor measurement and one and one-half days to criterion measurement.

While the same predictor battery was used for all the MOS, the criterion measures used for Batch A MOS were different than those used for MOS in Batch Z. The major distinction is that the MOS-specific job performance and job knowledge measures were not developed for the 10 MOS in Batch Z. For these jobs only Army-wide measures and the training achievement tests were administered.

As noted previously, the concurrent data were collected by traveling teams of contractor personnel, assisted by support staff from the specific Army post being visited.

DATA COLLECTION TEAM COMPOSITION AND TRAINING

Team Composition

Each data collection team was composed of a test site manager (TSM) and six or seven project staff members who were responsible for administering tests and rating scales. The teams were made up of a combination of regular project staff and individuals (e.g., graduate students) specifically recruited for the data collection effort. The test site manager was an "old hand" who had participated extensively in the field tests. This team was assisted by eight NCO scorers (for the hands-on tests), one company-grade officer POC, and up to five NCO support personnel, all recruited from the post.

Team Training

The project data collection teams were given 3 days of training at a central location (Alexandria, VA). During this period, Project A was explained in detail, including both operational and scientific objectives. After discussing the logistics of how the team would operate (transportation, meals, etc.), staff members presented a detailed explanation of the procedures for data entry from the field to the computer file. Every effort was made at the outset to reduce data entry errors by training team members in correct recording of responses and correct identification of answer sheets and diskettes.

Next, each predictor and criterion measure was examined and explained. The trainees took each predictor test, worked through samples of the knowledge tests, and role played the part of a rater. Considerable time was spent on the nature of the rating scales, rating errors, rater training, and the procedures to be used for administering the ratings. All administrative

Table 2.3

Summary of Predictor Measures Used in Concurrent Validation: The Trial Battery

Name	Number of Items
COGNITIVE PAPER-AND-PENCIL TESTS	
Reasoning Test (Induction - Figural Reasoning)	30
Object Rotation Test (Spatial Visualization - Rotation)	90
Orientation Test (Spatial Orientation)	24
Maze Test (Spatial Visualization - Scanning)	24
Map Test (Spatial Orientation)	20
Assembling Objects Test (Spatial Visualization - Rotation)	32
COMPUTER-ADMINISTERED TESTS	
Simple Reaction Time (Processing efficiency)	15
Choice Reaction Time (Processing efficiency)	30
Memory Test (Short-term memory)	36
Target Tracking Test 1 (Psychomotor precision)	18
Perceptual Speed and Accuracy Test (Perceptual speed and accuracy)	36
Target Tracking Test 2 (Two-hand coordination)	18
Number Memory Test (Number operations)	28
Cannon Shoot Test (Movement judgment)	36
Identification Test (Perceptual speed and accuracy)	36
Target Shoot Test (Psychomotor precision)	30
NON-COGNITIVE PAPER-AND-PENCIL INVENTORIES	
Assessment of Background and Life Experiences (ABLE)	209
Adjustment	
Dependability	
Achievement	
Physical Condition	
Leadership	
Locus of Control	
Agreeableness/Likability	
Army Vocational Interest Career Examination (AVOICE)	176
Realistic Interests	
Conventional Interests	
Social Interests	
Enterprising Interests	
Artistic Interests	

Table 2.4

Summary of Criterion Measures Used in Batch A and Batch Z Concurrent Validation Samples

---

Performance Measures Common to Batch A and Batch Z

- Army-wide rating scales (all obtained from both supervisors and peers).
    - Ten behaviorally anchored rating scales (BARS) designed to measure factors of non-job-specific performance.
    - Single scale rating of overall effectiveness.
    - Single scale rating of NCO potential.
- Combat Prediction scale containing 40 items.
- Paper-and-pencil tests of training achievement developed for each of the 19 MOS (130-210 items each).
- Personnel File Information form developed to gather objective archival records data (awards and letters, rifle marksmanship scores, physical training scores, etc.).

Performance Measures for Batch A Only

- Job sample (hands-on) tests of MOS-specific task proficiency.
    - Individual is tested on each of 15 major job tasks in an MOS.
- Paper-and-pencil job knowledge tests designed to measure task-specific job knowledge.
    - Individual is scored on 150 to 200 multiple-choice items representing 30 major job tasks. Ten to 15 of the tasks were also measured hands-on.
- Rating scale measures of specific task performance on the 15 tasks also measured with the knowledge tests. Most of the rated tasks were also included in the hands-on measures.
- MOS-specific behaviorally anchored rating scales (BARS). From six to 12 BARS were developed for each MOS to represent the major factors that constitute job-specific technical and task proficiency.

Performance Measures for Batch Z Only

- Additional Army-wide rating scales (all obtained from both supervisors and peers).
    - Ratings of performance on 11 common tasks (e.g., basic first aid).
    - Single scale rating on performance of specific job duties.

Auxiliary Measures Included in Criterion Battery

- A Job History Questionnaire which asks for information about frequency and recency of performance of the MOS-specific tasks.
- Army Work Environment Questionnaire - 53 items assessing situational/environmental characteristics, plus 46 items dealing with leadership.
- Measurement Method Rating obtained from all participants at the end of the final testing session.

---

manuals, which had been prepared in advance, were studied and pilot tested, roleplaying exercises were conducted, and hands-on instruction for maintaining the computerized test equipment was given.

The intent was that by the end of the 3-day session each team member would (a) be thoroughly familiar with all predictor tests and performance measures, (b) understand the goals of the data collection and the procedure for avoiding negative critical incidents, (c) have had an opportunity to practice administering the instruments and to receive feedback, and (d) be committed to making the data collection as error-free as possible.

Hands-On Scorer Training

As noted above, eight NCO scorers were required to administer and score the hands-on tests. They were recruited and trained at each post, using procedures very similar to those used in the criterion field tests (see Pulakos & Borman, 1986a). The purpose of the training was to develop high agreement among the scorers as to the precise responses that would be scored as GO or NO-GO on each step. The training required one full day and began with a thorough briefing on Project A. The scorers had the opportunity to take the tests themselves, checked out specified equipment, and underwent multiple practice trials in scoring each task, with feedback from the project staff.

DATA COLLECTION PROCEDURE

The Concurrent Validation administration schedule for a typical site (Fort Stewart, Georgia) is shown in Figure 2.2. The first day (22 Jul 85) was devoted to equipment and classroom set-up, general orientation to the data collection environment, and a training and orientation session for the post POC and the NCO support personnel.

On the first day of actual data collection (23 Jul 85), 30 MOS 12B soldiers arrived at the test site at 0745. The 30 soldiers were divided randomly into two groups of 15 soldiers each, identified as Group 1 or 2. Each group was directed to the appropriate area to begin the administration for that group. They rotated under the direction of the test site manager through the appropriate block of tests according to the schedule.

For soldiers in a Batch Z MOS, like 12B, the procedure took one day. For soldiers in a Batch A MOS, like 91A, the procedure was similar but took two days to rotate the soldiers through the appropriate blocks, as shown in the 6-7 August schedule at Fort Stewart. The measures administered in each block are shown in Figure 2.2.

Fort Stewart, GA
Concurrent Validation
22 July - 30 August 1985

	SL 10-20 Soldiers for 2 Days* (Batch A)				SL 10-20 Soldiers for 1 Day* (Batch Z)	
Groups of 15	1	2	3	4	1	2
22 Jul AM	Training/Orientation for Data Collection					
PM	Training/Orientation for Data Collection					
					30 12B Soldiers	
23 Jul AM					P	K/R
PM					R/K	P
					32 27E Soldiers	
24 Jul AM					K/R	P
PM					P	R/K
					30 55B Soldiers	
25 Jul AM					P	K/R
PM					R/K	P
					30 55B Soldiers	
26 Jul AM					K/R	P
PM					P	R/K
					30 76W Soldiers	
29 Jul AM					P	K/R
PM					R/K	P
					30 76W Soldiers	
30 Jul AM					K/R	P
					P	R/K
					30 94B Soldiers	
31 Jul AM					P	K/R
PM					R/K	P

Figure 2.2. Sample schedule for Concurrent Validation administration.
(Page 1 of 4)

		SL 10-20 Soldiers for 2 Days* (Batch A)				SL 10-20 Soldiers for 1 Day* (Batch Z)	
Groups of 15		1	2	3	4	1	2
						30 16S Soldiers	
1 Aug	AM					K/R	P
	PM					P	R/K
						15 16S & 14 51B Soldiers	
2 Aug	AM					P	K/R
	PM					R/K	P
						30 51B Soldiers	
5 Aug	AM	(Train 8 91A Scorers)				P	K/R
	PM			- -		R/K	P
		45 91A Soldiers				15 54E Soldiers	
6 Aug	AM	P	K_3R_1	HO		K/R	
	PM	HO	R_2K_5	R_1K_5		P	
7 Aug	AM	(Train 8 11B Scorers)					
	AM	R_1K_3	HO	P			
	PM	K_5R_2	P	R_2K_3			
		45 11B Soldiers				15 54E Soldiers	
8 Aug	AM	P	K_3R_1	HO		K/R	
	PM	HO	R_2K_5	R_1K_5		P	
9 Aug	AM	(Train 8 13B Scorers)					
	AM	R_1K_3	HO	P			
	PM	K_5R_2	P	R_2K_3			
		45 13B Soldiers				15 54E Soldiers	
12 Aug	AM	K_3R_1	HO	K_3R_1		P	
	PM	HO	P	R_2K_5		R/K	

Figure 2.2. Sample schedule for Concurrent Validation administration.
(Page 2 of 4)

		SL 10-20 Soldiers for 2 Days* (Batch A)				SL 10-20 Soldiers for 1 Day* (Batch Z)	
Groups of 15		1	2	3	4	1	2
		45 13B Soldiers				15 54E Soldiers	
13 Aug	AM	(Train 8 63B Scorers)					
	AM	P	R_1K_5	HO		K/R	
	PM	R_2K_5	R_2K_3	P		P	
		30 63B Soldiers					
14 Aug	AM	P	HO				
	PM	R/K	P				
						30 67N Soldiers	
15 Aug	AM	(Train 8 95B Scorers)				P	K/R
	AM	K_3R_1	R_1K_5			R/K	P
	PM	R_2K_5	R_2K_3				
		45 95B Soldiers					
16 Aug	AM	K_3R_1	HO	K_3R_1			
	PM	HO	P	R_2K_5			
19 Aug	AM	(Train 8 71L Scorers)					
	AM	P	R_1K_5	HO			
	PM	R_2K_5	R_2K_3	P			
		45 71L Soldiers				15 67N Soldiers	
20 Aug	AM	K_3R_1	HO	K_3R_1		P	
	PM	HO	P	R_2K_5		R/K	
21 Aug	AM	(Train 8 31C Scorers)					
	AM	P	R_1K_5	HO			
	PM	R_2K_5	R_2K_3	P			
		30 31C Soldiers				30 76Y Soldiers	
22 Aug	AM	HO	K_3R_1			K/R	P
	PM	R_1K_5	HO			P	R/K

Figure 2.2. Sample schedule for Concurrent Validation administration.
(Page 3 of 4)

Groups of 15		SL 10-20 Soldiers for 2 Days* (Batch A) 1	2	3	4	SL 10-20 Soldiers for 1 Day* (Batch Z) 1	2
23 Aug	AM	(Train 8 64C Scorers)					
	AM	R_2K_3	P				
	PM	P	R_2K_5				
		45 64C Scorers				15 76Y Soldiers	
26 Aug	AM	K_5R_1	K_5R_1	HO		P	
	PM	P	HO	R_1K_5		R/K	
27 Aug	AM	(Train 8 19E Scorers)					
	AM	HO	K_3R_2	P			
	PM	K_3R_2	P	R_2K_3			
		60 19E Soldiers					
28 Aug	AM	K_5R_1	P	K_3R_1	HO		
	PM	HO	R_1K_3	P	K_3R_1		
29 Aug	AM	K_3R_2	HO	R_2K_5	P		
	PM	P	K_5R_2	HO	R_2K_5		
30 Aug	AM	Make-up Day					
	PM	--					

---

*Legend:

R - Rating Scales
- R_1 Batch A (Army-Wide, MOS BARS, Job History)
- R_2 Batch A (Combat Prediction, Work Questionnaire, Personnel File Information)
- R Batch Z (Army-Wide, Overall Performance, Common Tasks, Combat Prediction, Work Questionnaire, Personnel File Information)

K - Knowledge Tests
- K_3 Batch A Training Achievement Tests (School Knowledge)
- K_5 Batch A MOS Task-Based Tests (Job Knowledge)
- K Batch Z Training Achievement Tests (School Knowledge)

P - Predictor Tests
HO - Hands-On Tests

In addition, at the end of their final session, all soldiers filled out the Measurement Method Rating (MMR).

Figure 2.2. Sample schedule for Concurrent Validation administration.
(Page 4 of 4)

CONCURRENT VALIDATION ANALYSIS

The basic analytic steps for the Concurrent Validation data are outlined below. The overall goal is to move systematically from the raw data, which consist of thousands of elements of information on each individual, to estimates of selection validity, differential validity, and selection/classification utility.

General Steps

The overall analysis plan consists of the following steps:

1. Prepare and edit individual data files.

2. Determine basic scores for the predictor variables.

3. Determine basic scores for the criterion variables.

4. Describe the latent structure of the predictor and criterion covariance matrixes.

5. Determine how well each predictor construct predicts each criterion factor (for each MOS).

6. Determine incremental validities (if any) of new predictors over ASVAB for each criterion factor within each MOS.

Data Preparation

For initial processing, the data from the field were divided into the following groups:

Predictor Measures -

- o Computer Tests - diskettes sent to project staff for processing.

- o Paper-and-Pencil Tests - booklets sent to vendor for scanning.

Criterion Measures -

- o Hands-on Measures - score sheets sent to project staff for keypunching.

- o Ratings, Knowledge Tests, Background, Job History, Work Questionnaire, Method Measurement, Personnel File Information - sent together to vendor for scanning.

The Roster Control File was merged with the most recent Enlisted Master Files extracts for the FY83/84 cohort and with Applicant/Accession files. Unmatched cases were further checked for incorrect identifiers.

Score Generation

While the data were still separated into the different types, initial score variables were generated:

- For the paper-and-pencil tests, number correct and number omitted scores were used. A missing data screen identified any score where more than 10 percent of the component items were omitted.

- For the non-cognitive predictor tests, scale scores established during the field tests were revised to reflect scoring changes suggested by item analyses as well as subsequent outlier analyses. A missing data screen identified any score where more than 10 percent of the component items were omitted.

- For the computer-administered tests, response time, error, and other derivative scores were generated as per the guidance from the field test results.

- For the hands-on measures, the percentage of steps passed was computed for each task. If more than 10 percent of the steps were not scored for a given soldier, the task score was identified as missing.

- For the rating data, adjusted ratee means were computed for each rating scale, as was done in the field tests.

- For the administrative measures, scores were recorded from the Personal File Information Form and computed from data obtained from the Accessions file.

Missing Values

Because extensive multivariate analyses requiring complete data were to be performed, the treatment of missing values was an important concern, much more so than was the case with the field test data. Typically, either examinee means or variable means are substituted for missing values. For these data, a statistical procedure known as PROC IMPUTE (see Chapter 3) was used to derive proxy values for missing scale scores, and for missing step scores in the hands-on analyses.

The PROC IMPUTE procedure essentially substitutes for the missing variable a value observed for a respondent who was very similar to the examinee. This procedure has been shown to be significantly better than ordinary least squares (OLS) regression procedures (e.g., BMDPAM) in reproducing correlation and variance estimates, as the regression approaches tend to underestimate variances and to spuriously inflate correlations.

Predictor Score Analyses

After data preparation, basic item analyses, and the initial score generation, the principal objectives for the predictor analyses were to generate the basic summary scores that would enter the initial prediction equation for each MOS. The basic steps were as follows:

1. Using the initial scores, conduct item/scale score analyses.

2. Compute scale reliabilities and descriptive statistics.

3. Develop predictor construct scores via factor analysis.

4. Estimate predictor factor (construct) scores via a simple weighted sum.

Criterion Score Analyses

After data preparation had been completed, the objectives for the criterion analyses were to identify an array of basic criterion variables (i.e., scores), investigate the latent structure of those variables, and determine the criterion construct scores. The following steps were taken:

1. A final set of item by a priori scale analyses was used to identify faulty or misplaced items. At this point the number of criterion variables was still too large to enter into an intercorrelation matrix. For example, the job knowledge test still contained 30 task scores.

2. A more manageable set of basic criterion scores was obtained by factoring/clustering rating scales, administrative measures, hands-on test steps, and knowledge test items. Exploratory factor analysis was used to reduce the individual rating scales to clusters of scales that could be averaged. For the hands-on and knowledge tests, items were clustered via expert judgment sorts.

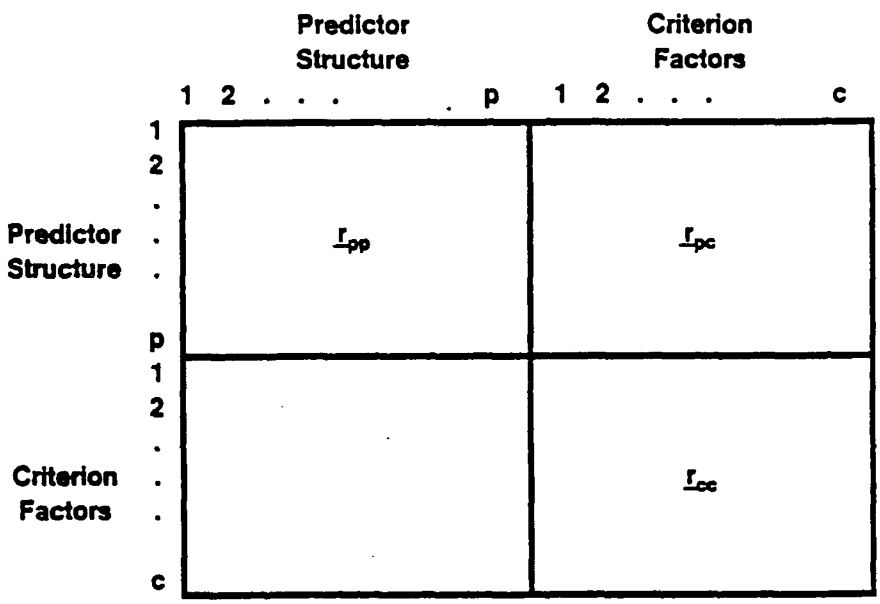
3. After Step 2 yielded a basic array of criterion scores, an intercorrelation matrix was calculated for each MOS. Exploratory factor analyses were used to generate hypotheses about the latent structure of the criterion space.

4. The "theories" about the criterion space generated in Step 3 were subjected to confirmatory analyses in an attempt to make a reasonable choice about the best-fitting model for the total domain of job performance for each MOS.

5. After the variables that comprise each criterion factor (construct) were identified, factor scores were generated by computing a simple sum.

Predictor/Criterion Interrelationships

After the above steps were carried out, the basic variables and the best-fitting model for both the predictors and the performance measures had been identified. They provide the variables to be used for establishing the selection/classification validity of the new predictor battery and for determining differential validity across criterion constructs, across jobs, and across subgroups. The basic steps in the validity analysis are summarized in Figure 2.3.



NOTE: There are three principal categories of analyses. First, the basic predictor scores must be generated, based on the predictor covariance structure ($\underline{r}_{pp}$). Second, a similar analysis must be carried out on the criterion covariance ($\underline{r}_{cc}$). Third, the validity coefficients themselves ($\underline{r}_{pc}$) can be analyzed.

A. Within MOS:

1. Compute "best" prediction equation for each criterion factor.
2. Compute best prediction equation for overall composite score.
3. Determine loss of predictability as number of equations is reduced from (1) to (2).
4. Determine incremental validity of Project A measures (over ASVAB).

B. Between MOS:

1. Determine generalizability of each performance factor's prediction equation across MOS.
2. Determine generalizability of composite prediction equations across MOS.
3. Determine generalizability of incremental validity across MOS.

Figure 2.3. Principal categories of analysis.

CONCLUDING COMMENTS

This chapter provided a brief outline of the nature of the Concurrent Validation sample, the predictor and criterion measures that were used, the data collection procedures, and the basic data analytic steps that were planned and undertaken. The results of the analytic steps that have been completed are summarized in the ensuing chapters.

Chapter 3

TREATMENT OF MISSING DATA[1]

The procedures for collecting Concurrent Validation data were subjected to extensive pilot and field tests. The data collection teams were carefully trained and were supervised by senior staff. The quality and completeness of the data collected attest to the thoroughness of these procedures. However, notwithstanding our best efforts, the final data were to some extent incomplete. The purpose of this chapter is to describe the amount of missing data in the project CV data base, the problems posed by incomplete data, and the steps taken to overcome those problems.

REASONS FOR INCOMPLETE DATA

Figure 3.1 lists some of the chief reasons for missing CV data. Most of the reasons are self-explanatory, but examples involving the hands-on tests may help to illustrate some of the problems we encountered.

At Fort Hood, Texas, we were testing Armor Crewmen when a spring in the breech block of one of the howitzers failed. On that particular occasion, we had arranged for a back-up howitzer. Consequently, we did not lose any data.

During hands-on testing of Infantrymen at Fort Benning, Georgia, we were not so fortunate. The afternoon started bright and sunny, so we decided to administer the tests at our primary testing site, near a meandering creek, rather than at our back-up bad weather site. The weather in central Georgia is notoriously fickle on summer afternoons, and a short time later we were caught in a deluge. The creek rose. Everyone was up to their shins in water, and our test administrators were scrambling madly, trying to protect their equipment and scoresheets from the driving rain. Unfortunately, one test administrator simply was not quick enough. As he and the hands-on test site manager watched, two scoresheets began to float away. Before they could be reached, they were sucked into the creek and carried swiftly downstream. The thunderstorm abated a short while later, but valuable time had been lost and it was not possible to move all of the subjects through all of the test stations before the soldiers' work day ended. As a result, the session ended with quite a bit of data missing.

Two other problems encountered during hands-on testing were equipment variation and scoresheet errors. In most cases, we were able to make allowances for equipment variation by developing parallel forms of a test.

---

[1] The material in this chapter is drawn from Project A Concurrent Validation: Treatment of missing data, by Lauress L. Wise, Jeffrey J. McHenry, and Winnie Y. Young, ARI Working Paper RS-WP-86-08, 1986.

HANDS-ON DATA

- o Anticipated variation in equipment
- o Unanticipated variation in equipment
- o Soldiers not available for part or all of scheduled time
- o Equipment breakdown or nonavailability
- o Conditions preventing testing of some soldiers on some tasks
- o Scorer or scoresheet errors

RATING DATA

- o No suitable raters available
- o Soldier does not perform some kinds of tasks
- o Rater not following instructions

KNOWLEDGE TEST

- o Soldiers not available for part or all of scheduled time
- o Soldiers exceptionally slow in taking test
- o Soldiers not following instructions

ADMINISTRATIVE MEASURES

- o Soldier did not know scores for the self-report form
- o Information not available on accession file

Figure 3.1. Some reasons for incomplete Concurrent Validation data.

Often, this involved omitting certain steps that were irrelevant for one of the equipment models; in other cases, parallel sets of steps were developed. We tried to make this procedure clear on our scoresheets and in scorer training, but on a few occasions we were not successful. For example, in one hands-on test for Single Channel Radio Operators, the scoresheets for one task included a set of steps to be scored for one type of equipment and a different set to be scored for another type of equipment; no subject should have had scores for all of the steps. Nevertheless, two cases had data for both sets of steps, creating a unique problem of "too much" data rather than missing data. In several other instances, a scorer had trouble understanding some of the directions on the scoresheets and left one or more steps unscored.

Finally, a problem that plagued us throughout our testing was that subjects often had other commitments or were called away in the midst of tests. A subject might get halfway through a test, then have to leave for a dentist appointment that had been scheduled two or three months previously. These unavoidable absences doubtless caused more missing data than any other factor listed in Figure 3.1.

AMOUNT OF MISSING DATA

For any given instrument, data may be either partially missing (i.e., the soldier failed to complete some items or steps) or totally missing (i.e., the soldier was not available for a testing session). Moreover, if data are partially missing, amounts missing may be relatively small or relatively large.

Table 3.1 shows the extent of missing data for the school knowledge (SK) tests. There were only a few instances (1%) in which a soldier failed to take the test at all. There were also very few soldiers (1%) who had relatively large amounts of missing data. There were, however, a significant number of cases (16%) in which a small number of items had been omitted.

Table 3.1 also shows small differences between the Batch A and Batch Z MOS in the proportion of soldiers not tested at all. For all but one of the Batch A MOS the percentage not tested is above 1 percent, while the percentage not tested is below 1 percent for all but one of the Batch Z MOS. This difference is a direct consequence of the fact that all of the Batch Z testing took place in a single day while the Batch A testing required two full days of a soldier's time.

Table 3.2 shows the extent of missing data for the job knowledge (JK) tests. (Subjects in Batch Z MOS did not complete job knowledge or hands-on tests.) Again, there were very few instances (1%) where soldiers were not tested at all. The proportions of soldiers with relatively small (20%) and relatively large (3%) amounts of missing data are slightly higher than for the SK tests, but are generally quite comparable.

Table 3.3 shows the extent of missing data for the hands-on (HO) tests. The number of soldiers not tested was again small (1.8%). The number of soldiers with at least some data missing was, in many cases, very large. For the most part, these instances were due to equipment variation or failure.

Table 3.4 shows the extent of missing data for the rating measures. A scale or instrument was counted as present if at least one peer or at least one supervisor provided a rating. With the exception of the job task ratings (JTR), all the completion rates were quite high. The JTR scales provided a "cannot rate" option that was counted as missing, and this option accounts for most instances of partially missing data. Tables 3.5 and 3.6 show the same information for supervisors and peers separately. The percentage of soldiers with no ratings was quite a bit higher (8.4%) than when supervisor and peer ratings were combined, because no appropriate peer or no appropriate supervisor was available in many instances.

Table 3.1

Number and Percentage of Cases With Incomplete School Knowledge Data for Each MOS

MOS	No Data Missing		Less Than 10% Missing		More Than 10% Missing		All Data Missing		Total Number
	No.	%	No.	%	No.	%	No.	%	
BATCH A									
11B	504	86.0	88	12.5	2	0.3	8	1.1	702
13B	538	80.7	110	16.5	5	0.8	14	2.1	687
19E	403	80.1	88	17.5	4	0.8	8	1.6	503
31C	314	85.8	40	10.9	1	0.3	11	3.0	365
63B	536	84.1	81	12.7	10	1.6	10	1.6	637
64C	583	85.0	93	13.6	3	0.4	7	1.0	686
71L	458	89.1	41	8.0	2	0.4	13	2.5	514
91A	423	84.4	61	12.2	2	0.4	15	3.0	501
96B	583	84.2	100	14.4	3	0.4	6	0.8	692
BATCH B									
12B	569	80.8	124	17.6	5	0.7	6	0.8	704
16S	402	85.5	67	14.3	0	0.0	1	0.2	407
27E	111	75.5	34	23.1	2	1.4	0	0.0	147
51B	88	81.5	14	13.0	5	4.6	1	0.9	108
54E	350	80.6	80	18.4	2	0.5	2	0.5	434
55B	209	71.8	65	22.3	15	5.2	2	0.7	291
67N	155	56.2	116	42.0	5	1.8	0	0.0	276
76W	388	79.2	90	18.4	10	2.0	2	0.4	490
76Y	487	77.3	119	18.9	19	3.0	5	0.8	630
94B	474	77.4	116	19.0	14	2.3	8	1.3	612
Total	7,675		1,527		109		119		9,430

Table 3.2

Number and Percentage of Cases With Incomplete Job Knowledge Data for Each Batch A MOS

MOS	No Data Missing No.	No Data Missing %	Less Than 10% Missing No.	Less Than 10% Missing %	More Than 10% Missing No.	More Than 10% Missing %	All Data Missing No.	All Data Missing %	Total Number
11B	506	72.1	180	25.6	7	1.0	9	1.3	702
13B	460	69.0	180	27.0	17	2.6	10	1.5	667
19E	350	69.6	115	22.9	30	6.0	8	1.6	503
31C	304	83.1	24	6.6	31	8.5	7	1.9	366
63B	481	75.5	120	18.8	26	4.1	10	1.6	637
64C	533	77.7	141	20.6	5	0.7	7	1.0	686
71L	395	76.8	107	20.8	6	1.2	6	1.2	514
91A	428	85.4	59	11.8	9	1.8	5	1.0	501
95B	595	86.0	74	10.7	21	3.0	2	0.3	692
Total	4,052		1,000		152		64		5,268

Table 3.3

Number and Percentage of Cases With Incomplete Hands-On Data for Each Batch A MOS

MOS	No Data Missing No.	No Data Missing %	Less Than 10% Missing No.	Less Than 10% Missing %	More Than 10% Missing No.	More Than 10% Missing %	All Data Missing No.	All Data Missing %	Total Number
11B	188	26.8	471	67.1	30	4.3	13	1.8	702
13B	184	27.6	351	52.6	114	17.1	18	2.7	667
19E	341	67.8	131	26.0	18	3.6	13	2.6	503
31C	2	0.6	228	62.3	125	34.2	11	3.0	366
63B	135	21.2	380	59.6	106	16.6	16	2.5	637
64C	132	19.2	433	63.1	112	16.3	9	1.3	686
71L	244	47.5	218	42.4	46	9.0	6	1.2	514
91A	346	69.1	145	28.9	5	1.0	5	1.0	501
95B	326	47.1	308	44.5	56	8.1	2	0.3	692
Total	1,898		2,665		612		93		5,268

Table 3.4

Percentage of Cases With Missing Data by Rating Instrument, Using Combined Supervisor and Peer Ratings (All MOS: N = 9,430)

Instrument	No Data Missing	1-10% Data Missing	More Than 10% Missing	All Data Missing
Army-Wide BARS[a]	98.3	0.2	0.0	1.5
MOS-Specific BARS	97.0	0.3	0.9	1.8
Task Ratings	66.2	11.2	20.1	2.4
Combat Prediction	98.3	0.1	0.1	1.5
All Instruments	66.0	18.7	3.8	1.5

[a]BARS = Behaviorally Anchored Rating Scale.

Table 3.5

Percentage of Cases With Missing Data by Rating Instrument, for Supervisor Ratings Only (All MOS: N = 9,430)

Instrument	No Data Missing	1-10% Data Missing	More Than 10% Missing	All Data Missing
Army-Wide BARS[a]	90.3	0.9	0.3	8.5
MOS-Specific BARS	82.7	2.3	5.3	9.8
Task Ratings	30.2	13.5	45.3	10.9
Combat Prediction	89.4	1.8	0.2	8.6
All Instruments	29.2	50.0	12.3	8.4

[a]BARS = Behaviorally Anchored Rating Scale.

Table 3.6

Percentage of Cases With Missing Data by Rating Instrument, for Peer Ratings Only (All MOS: N = 9,430)

Instrument	No Data Missing	1-10% Missing	More than 10% Missing	All Data Missing
Army-Wide BARS[a]	91.0	0.4	0.2	8.4
MOS-Specific BARS	88.9	0.5	1.3	9.3
Task Ratings	48.1	11.0	30.0	11.0
Combat Prediction	90.4	0.9	0.2	8.6
All Instruments	47.5	34.2	9.9	8.4

[a]BARS = Behaviorally Anchored Rating Scale.

Table 3.7 shows the amount of missing job task ratings (Batch A MOS) and common task ratings (CTR) (Batch Z MOS) data by MOS. There was considerable variation across MOS. For some MOS (e.g., MOS 12B, Combat Engineer; MOS 16S, MANPADS Crewman), data completeness levels were very high. However, for MOS where soldiers tend to work in isolation from other soldiers in their MOS and tend to perform only a subset of the tasks rated, the incidence of missing data was significantly higher. The best example is MOS 71L, Administrative Specialist, in which ratings data were complete for only 24 percent of the subjects.

From the results presented thus far, it might be tempting to conclude that, except for the JTR/CTR data, missing data were not a significant problem in analyses of the Project A CV data. Figure 3.2 indicates that this is not the case. The figure shows the number of Batch A soldiers with different patterns of complete and missing data across the four performance measurement methods. Fewer than 15 percent of the cases in the entire sample have complete data for all four methods. If the ratings data are set aside, there are still fewer than 25 percent of the subjects with complete HO, JK, and SK data. Similarly, ignoring the HO data still leaves about 42 percent of the CV subjects with complete data on the remaining measures. Whether or not the sample of soldiers with complete data is representative of the target population, the sheer loss of statistical power associated with such a reduction in sample size would be unacceptable.

Since the administrative measure did not include a large number of component parts, only single scores were obtained for each soldier on these measures. Physical Readiness scores were missing for 10 to 15 percent of the examinees. Similarly, Promotion Rate Deviation scores were missing for about 15 percent, primarily because of problems in retrieving accession file information needed to compute time in service. Only a small percentage of scores were missing on the other administrative measures.

Table 3.7

Percentage[a] of Cases With Missing Task Ratings by MOS, Using Combined Supervisor and Peer Ratings

MOS	No Data Missing	1-10% Missing	More than 10% Missing	All Data Missing	Total N
Batch A					
11B	71.5	14.4	12.4	1.7	702
13B	75.4	6.4	17.5	0.6	667
19E	68.8	14.5	16.3	0.4	503
31C	56.3	16.4	24.6	2.7	366
63B	63.3	11.0	22.9	2.8	637
64C	60.5	9.9	27.0	2.6	686
71L	23.9	18.3	53.9	3.9	514
91A	60.7	13.2	25.4	0.8	501
95B	70.4	11.8	17.6	0.1	692
Batch B					
12B	93.3	3.1	2.4	1.1	704
16S	91.5	5.3	3.2	0.0	470
27E	74.2	6.8	18.4	0.7	147
51B	84.3	5.6	8.3	1.8	108
54E	73.7	12.2	10.4	3.7	434
55B	69.4	12.7	16.2	1.7	291
67N	62.3	13.8	22.8	1.1	276
76W	61.2	14.5	20.2	4.1	490
76Y	49.0	11.6	31.4	7.9	630
94B	59.2	10.5	25.2	5.2	612
All MOS	66.2	11.2	20.2	2.4	9,430

[a]Percentages do not add to 100 due to rounding error.

	Complete SK JK	Comp SK Miss JK	Miss SK Comp JK	Missing SK JK	TOTAL
Complete HO & RA	772 14.6%	189 3.6%	122 2.3%	58 1.1%	1,141 21.7%
Comp HO Miss RA	526 10.0%	130 2.5%	72 1.4%	29 0.6%	757 14.4%
Miss HO Comp RA	1,436 27.3%	364 6.9%	215 4.1%	125 2.4%	2,140 40.6%
Missing HO & RA	784 14.9%	241 4.6%	125 2.4%	80 1.5%	1,230 23.4%
TOTAL	3,518 66.8%	924 17.5%	534 10.1%	292 5.5%	5,268 100.0%

NOTE: HO = Hands-On; SK = School Knowledge; JK = Job Knowledge; RA = Rating

Figure 3.2. Number and percentage of cases with complete data for each combination of criterion instruments: Batch A.

TREATMENT OF MISSING DATA

The processing of missing data was approached in two stages. In the first stage, we focused on one instrument at a time and dealt with only those subjects for whom a small amount of data was missing on the instrument under consideration. In the second stage, we formulated procedures for dealing with subjects for whom all or a high percentage of the data were missing on a given instrument.

Stage I: Missing Data Within Each Instrument

Amount of Missing Data Permitted. The first step was to decide how much missing data was too much. The frequency of missing data formed somewhat of a bimodal distribution. Most soldiers had only a few missing steps, items, or scales, while a smaller number of soldiers had all or nearly all elements missing. For each instrument, we made a judgment about where to conservatively set the dividing line. For cases with minimal missing data, we would take steps to fill in missing values so as to be able to compute performance scores on that instrument. For cases with significant amounts of missing data, we would not attempt to compute performance scores for the instrument in question.

In general, we sought to retain 90-95 percent of the soldiers tested in each MOS, but to eliminate cases with more than 10 percent elements missing. For the written tests (JK and SK), a cutoff of 10 percent missing would still retain well over 95 percent of the subjects in each MOS. For hands-on and each of the rating instruments, a slightly more liberal cutoff of 15 percent missing was chosen as the best balance between the desire to retain most of the cases and the desire to limit strongly the number of values that must be imputed to achieve complete data. For the hands-on data we adopted a two-stage rule. For each task tested, a task score was generated only if no more than 15 percent of the steps were missing. We then computed overall hands-on scores only if no more than three task scores (no more than four task scores for MOS 31C and MOS 63B, where we had relatively small samples) were missing.

Elimination of Random Responders. In consultation with our advisory groups, we developed rules for identifying and eliminating "unreliable" or random responders on each instrument.

For the written tests, a random response index was defined as the correlation between the item score (1 for correct and 0 for incorrect) and item difficulty (expressed as proportion of subjects who answered the item correctly). For most examinees this correlation was positive, since subjects tended to get the easier items correct and miss the more difficult items. For a few examinees this correlation was essentially zero, suggesting random responding. For these subjects, all of their responses for that particular instrument were classified as missing.

For the hands-on data, random responding was not a concern. The data sheets were filled out by trained (and monitored) NCOs and not by the examinees themselves. There was no indication that any subjects intentionally responded poorly or randomly in front of the NCO scorers. No screening for unreliable responses in the hands-on data was conducted.

For the rating data, our concern was for unreliable raters rather than unreliable examinees. Reliability indexes were constructed for each rater by comparing the ratings made by an individual with the average of all other raters' ratings of the same soldiers on the same scales. Both mean difference and correlational indexes were used in identifying outliers among the raters.

Establishment of Separate Tracks to Account for Equipment Differences. For several MOS, the hands-on scoring differed for different equipment. To achieve comparable scores across these equipment differences, the examinees were separated into "tracks" corresponding to the different variations in equipment. (For Military Police, for example, females used and were tested on a .38 caliber hand gun while males used and were tested on a .45 caliber hand gun.) Minimal differences were found between track samples on those tasks that were scored the same. Consequently, scores for each track were standardized to have a mean and standard deviation that matched the original overall mean for the score in question.

Number of Subjects Dropped for Missing Data or Unreliable Responses. Table 3.8 shows the number of cases deleted for the SK tests because of too much missing data and/or because of apparent random responding. Table 3.9 shows similar results for the JK tests, and Table 3.10 shows the number of cases deleted due to too much missing data on the HO tests.

Table 3.8

Number of Cases With School Knowledge Data Deleted Because of Too Much Missing Data or Random Response

MOS	Missing More than 10%	Random Responses	Both	Total Dropped	Total N	Percent Dropped
Batch A						
11B	2	8	0	10	694	1.4
13B-S	3	10	0	13	536	2.4
13B-T	2	2	0	4	117	3.4
19E	4	6	0	10	495	2.0
31C	1	5	0	6	355	1.7
63B	10	5	0	15	627	2.4
64C	3	7	0	10	679	1.5
71L	3	5	0	8	501	1.6
91A	2	5	0	7	486	1.4
95B	4	9	0	13	687	1.9
Batch B						
12B	5	11	0	16	698	2.3
16S	0	1	0	1	469	0.2
27E	2	3	0	5	147	3.4
51B	4	0	1	5	107	4.7
54E	2	4	0	6	432	1.4
55B	15	1	0	16	289	5.5
67N	5	0	0	5	276	1.8
76W	9	5	1	15	488	3.1
76Y	19	10	0	29	625	4.6
94B	14	20	0	34	604	5.6

Table 3.9

Number of Cases With Job Knowledge Data Deleted Because of Too Much Missing Data or Random Response

MOS	Missing More than 10%	Random Responses	Both	Total Dropped	Total N	Percent Dropped
11B	9	6	0	15	693	2.2
13B	16	1	1	18	657	2.7
19E	29	6	1	36	495	7.3
31C	31	2	0	33	359	9.2
63B	26	4	1	31	627	4.9
64C	7	4	0	11	679	1.6
71L	6	1	0	7	508	1.4
91A	9	4	0	13	496	2.6
95B	22	3	0	25	690	3.6

Table 3.10

Number of Cases With Hands-On Data Deleted Because of Too Much Missing Data

MOS	Cases Deleted	Percent Dropped
11B	8	1.2
13B	37	5.7
19E	16	3.4
31C	14	4.1
63B	52	9.1
64C	37	5.5
71L	14	2.8
91A	0	0.0
95B	25	3.6

Elimination of unreliable raters did not result in the loss of rating data for any individual subjects. In all cases where raters were eliminated, other raters provided data on these subjects. (Where there were no other raters, the rater in question was not eliminated because there was no basis for estimating the reliability of the ratings.)

Imputation of Missing Values. After dropping subjects judged to be random responders and eliminating cases with too much missing data, values for the remaining missing data were imputed in such a way that total scores could be computed.

Several options for imputing scores were considered. The first was to compute the subject's mean on the variables that were present and then substitute this mean for each of the missing variables. This is equivalent to defining the total score as the mean of the values present. The problem with this approach was that items and steps differed considerably in difficulty. Systematic bias could be introduced by substituting the examinee's mean where data were missing.

The second option was to substitute the variable (item, step, scale) mean for all missing values on that variable. This option was rejected because it would reduce individual differences. Subjects performing at different levels should have different estimates for the missing items.

The option used to fill in missing values was a procedure that had been developed for the National Center for Education Statistics (now the Center for Education Statistics) known as PROC IMPUTE (Wise & McLaughlin, 1980). Several features of PROC IMPUTE made it preferable to other readily available options for filling in the missing CV values.

First, PROC IMPUTE uses regression estimates to predict missing values. Each missing value is predicted from other values for the subject in question so that individual differences are retained. The regression coefficient and intercept vary from item to item so that differences in item difficulty are also reflected in the predicted values.

Second, PROC IMPUTE adds a random variable with variance equal to the error of estimate for predicting the missing value. If such a random variable is not added, the imputed values are more highly correlated with values on other variables in comparison with nonimputed values.

Third, PROC IMPUTE employs a sequential strategy that maintains relationships between variables when more than one value is imputed for the same examinee. A two-stage approach is used, with the first variable imputed from nonmissing values. The second (and subsequent) variable(s) is imputed from the nonmissing values plus the imputed value for the first variable. After all initial imputations, values are reimputed in a second pass where all of the initially imputed values participate in the reimputation of each missing value.

Finally, PROC IMPUTE models nonlinear relationships between the predicted and actual values. If the actual values are discrete, PROC IMPUTE provides discrete values for the missing elements as well. Table 3.11 illustrates the final step in PROC IMPUTE. The predicted values were divided into six equal intervals to define predicted "levels". At each predicted level there were from 61 to 92 cases for whom actual technical skill ratings were available. The distribution (in percentages) of actual scores for each predicted level is shown. For each soldier with a missing technical skill rating, a predicted level is computed. (Actually, the program interpolates between predicted levels.) A uniformly distributed random number between 0 and 100 is generated and mapped onto the actual levels, using the cumulative distribution of actual scores for the predicted level. (Again the program actually interpolates between levels.) The actual level scores are then transformed back to the original units.

Table 3.11

Distribution of Technical Skill Ratings for Each Predicted Level

Predicted Level	Total No. of Cases	Percent at Each Actual Level: Level 1	Level 2	Level 3	Level 4	Level 5
1	67	15	57	18	0	0
2	61	0	21	77	2	0
3	92	0	7	65	28	0
4	89	0	0	40	59	1
5	92	0	0	8	91	1
6	71	0	0	3	52	45

PROC IMPUTE was used in all instances except one. For the written tests, a distinction was made between internal omissions (prior to the last item answered) and items that were not reached (omissions after the last item answered). For internal omissions, we assumed that the examinee did not know the answer and we substituted a score equal to the guessing rate (e.g., .2 for a 5-option item). If the actual proportion passing the item was lower than the guessing rate, the proportion passing was used instead of the guessing rate. We made no assumptions regarding items not reached since the examinee may not have had time to demonstrate knowledge of the item. Items not reached were imputed with PROC IMPUTE, as were all missing hands-on steps and rating scales.

Tables 3.12, 3.13, and 3.14 show the changes in summary statistics that resulted from the Stage I screening and imputations for three different MOS. Initial totals were computed using means of available data. The sample sizes dropped slightly due to screening out random responders and cases with too much data missing. Only small changes in means, standard deviations, reliabilities, and correlations resulted from the Stage I procedures. (Mean shifts for the first three scales should be compared against a standard deviation of 10.0, while the three rating factors were on a 7-point scale with a standard deviation of just under 1.0.)

Stage II: Missing Instruments

After cases were dropped or missing values were filled in on an instrument-by-instrument basis, the next decision was whether to estimate individual scores if only partial data were available for the individual. Again after considerable consultation, we decided on a 50 percent rule. An examinee had to have data on at least half of the instruments entering into a particular performance construct before we would estimate a score on the performance construct for that individual. Where 50 percent or fewer were missing, PROC IMPUTE was again used to fill in the missing pieces.

Table 3.15 shows the number of soldiers in each MOS who had missing values for each instrument after the imputations and screening were completed.

In most instances, the number of missing cases was quite small (1 or 2%). The chief exceptions were the two administrative measures noted earlier. Administrative measures had not been included in the Stage I imputation process because they do not include a large number of component parts. While Physical Readiness test scores were missing for 10 to 15 percent of the examinees, in most instances supervisor and peer ratings of physical fitness were available for these same examinees. Similarly, while Promotion Rate Deviation scores were missing for a significant number of cases (15%), for the most part variation in promotion rates among first-tour enlisted soldiers reflected instances where disciplinary problems led to delays in promotions; such delays were predicted fairly well from ratings of self-control and integrity and from the administrative index of disciplinary actions.

Tables 3.16, 3.17, and 3.18 show changes in summary statistics that resulted from Stage II imputations for the same three MOS illustrated previously. Again, only small changes resulted. There was a slight drop in hands-on means, because soldiers with missing hands-on scores tended to score well below average on other measures.

Table 3.12

Stage I Results: Changes in Statistical Characteristics of Summary Performance Measures From Pruning, Imputing[a], and Standardizing - MOS 11B, Infantryman

					Correlation					
Performance Measure[b]	N	Mean	SD	Reliability	SK	JK	HO	AWB 1	AWB 2	AWB 3
SK Total Score	-11	0.6	-0.7	-.01	.	.00	.05	.00	.02	-.01
JK Total Score	-15	0.1	-0.3	.01	.00	.	.05	.01	.01	.00
Hands-On Total	-4	-0.4	-0.9	.02	.05	.05	.	.05	.02	.06
AWB 1: Effort/Leadership	-1	-.01	.01	.02	.00	.00	.05	.	.00	.00
AWB 2: Personal Discipline	-1	-.01	.01	.02	.02	.01	.02	.00	.	.01
AWB 3: Personal Fitness/ Military Bearing	-1	.00	.01	.02	-.01	.00	.06	.00	.01	.

[a]Except ratings.

[b]SK = School knowledge; JK = Job knowledge; AWB = Army-Wide Behaviorally Anchored Rating Scale (BARS).

Table 3.13

Stage I Results: Changes in Statistical Characteristics of Summary Performance Measures From Pruning, Imputing[a], and Standardizing - MOS 63B, Light Wheel Vehicle Mechanic

					Correlation					
Performance Measure[b]	N	Mean	SD	Reliability	SK	JK	HO	AWB 1	AWB 2	AWB 3
SK Total Score	-15	0.8	-0.7	-.01	.	.02	.05	.02	.01	-.03
JK Total Score	-31	0.3	-0.4	.00	.02	.	.08	.00	.01	-.01
Hands-On Total	-40	-0.3	-1.2	-.03	.05	.08	.	-.02	-.01	-.08
AWB 1: Effort/Leadership	-2	-.01	.00	.02	-.02	.00	-.02	.	.00	.01
AWB 2: Personnel Discipline	-2	-.01	.00	.01	.00	.01	-.01	.00	.	.00
AWB 3: Military Bearing	-2	.01	.00	.02	-.03	-.01	-.06	.01	.00	.

[a]Except ratings.

[b]SK = School knowledge; JK = Job knowledge; AWB = Army-Wide Behaviorally Anchored Rating Scale (BARS).

Table 3.14

Stage I Results: Changes in Statistical Characteristics of Summary Performance Measures From Pruning, Imputing[a], and Standardizing - MOS 71L, Administrative Specialist

					Correlation					
Performance Measure[b]	N	Mean	SD	Reliability	SK	JK	HO	AWB 1	AWB 2	AWB 3
SK Total Score	-8	0.4	-0.3	-.01	.	.01	.03	.02	.02	.01
JK Total Score	-7	0.1	-0.1	.00	.01	.	.02	.01	.01	.00
Hands-On Total	6	-7.0	-0.8	.05	.03	.02	.	.00	.00	.01
AWB 1: Effort/Leadership	0	-.00	.00	.00	.02	.01	.01	.	.00	.00
AWB 2: Personal Discipline	0	.00	.00	.00	.02	.01	.00	.00	.	.00
AWB 3: Personal Fitness/ Military Bearing	0	.00	.00	.00	.01	.00	.01	.00	.00	.

[a]Except ratings.

[b]SK = School knowledge; JK = Job knowledge; AWB = Army-Wide Behaviorally Anchored Rating Scale (BARS).

Table 3.15

Number of Cases Missing for Each Instrument

	MOS								
Measure[a]	11B	13B	19E	31C	63B	64C	71L	91A	95B
Total N	702	667	503	366	637	686	514	501	692
Missing Hands-On	20	55	29	25	68	46	20	5	27
Missing Job Knowledge	24	29	44	40	41	18	13	18	29
Missing School Knowledge	18	28	18	17	25	17	21	22	18
Missing AW BARS	7	2	1	8	12	8	11	3	0
Missing MOS BARS	9	12	3	9	18	13	23	8	0
Missing Comb Pred	7	2	1	8	12	8	11	3	0
Missing A1: Awards	14	24	13	13	11	12	14	11	4
Missing A2: Phys Red.	63	93	53	30	80	81	60	59	57
Missing A4: Arts. 15	23	28	16	14	11	14	15	14	4
Missing A5: Prom Rt	109	143	83	62	97	86	79	61	84
Total Complete	512	406	335	241	411	486	355	374	513
Percentage Complete	72.9	60.6	66.6	65.9	64.5	70.9	69.1	74.7	74.1
Final Counts After Stage II Imputation									
Total N	693	656	490	356	615	675	506	492	686
Percentage of Original	98.7	98.4	97.4	97.3	96.6	98.5	98.4	98.2	99.1

[a]The administrative measures are:

A1 Awards and Certificates

A2 Physical Readiness

A4 Articles 15/Flag Actions

A5 Promotion Rate Deviation

Another administrative measure, the M16 qualification score, was carried in the analysis for a while as a unique variable. However, since it could not subsequently be demonstrated to possess any common variance, it was dropped from the analysis and was not used in the later scoring of criterion factors.

Table 3.16

Stage II Results: Changes in Statistical Characteristics of Summary Performance Measures From Pruning, Imputing[a], and Standardizing - MOS 11B, Infantryman

				Correlation					
Performance Measure[b]	N	Mean	SD	SK	JK	HO	AWB 1	AWB 2	AWB 3
SK Total Score	11	-.1	.1	.	.01	.02	.02	.01	.01
JK Total Score	16	-.5	.1	.01	.	.00	.01	.00	.01
Hands-On Total	15	-2.0	-.5	-.02	.00	.	.02	.01	.03
AWB 1: Effort/Leadership	0	.0	.01	.02	.01	.02	.	.00	.00
AWB 2: Personnel Discipline	6	-.01	.01	.01	.00	.01	.01	.	.00
AWB 3: Military Bearing	6	.00	.00	.01	.01	.03	.06	.00	.

[a]Except ratings.

[b]SK = School knowledge; JK = Job knowledge; AWB = Army-Wide Behaviorally Anchored Rating Scale (BARS).

Table 3.17

Stage II Results: Changes in Statistical Characteristics of Summary Performance Measures From Pruning, Imputing[a], and Standardizing - MOS 63B, Light Wheel Vehicle Mechanic

				Correlation					
Performance Measure[b]	N	Mean	SD	SK	JK	HO	AWB 1	AWB 2	AWB 3
SK Total Score	13	-.4	-.0	.	.00	.01	.01	.00	.00
JK Total Score	25	1.0	.0	.00	.	.02	.01	.00	.02
Hands-On Total	49	-.9	-.7	.01	.02	.	.00	.00	.05
AWB 1: Effort/Leadership	0	.00	.00	.00	.00	.00	.	.00	.00
AWB 2: Personnel Discipline	6	.00	.00	.00	.00	.00	.00	.	.00
AWB 3: Military Bearing	9	.00	.01	.00	.02	.05	.00	.00	.

a Except ratings.

b SK = School knowledge; JK = Job knowledge; AWB = Army-Wide Behaviorally Anchored Rating Scale (BARS).

Table 3.18

Stage II Results: Changes in Statistical Characteristics of Summary Performance Measures From Pruning, Imputing[a], and Standardizing - MOS 71L, Administrative Specialist

Performance Measure[b]	N	Mean	SD	Correlation SK	JK	HO	AWB 1	AWB 2	AWB 3
SK Total Score	11	.0	.0	.	.00	.03	.00	.00	.00
JK Total Score	6	.8	.0	.01	.	.01	.00	.02	.00
Hands-On Total	18	-4.3	-1.7	.03	.01	.	.01	.00	.02
AWB 1: Effort/Leadership	0	.00	.00	.00	.00	.01	.	.00	.00
AWB 2: Personnel Discipline	7	.00	.00	.00	.02	.00	.00	.	.01
AWB 3: Military Bearing	9	.00	.00	.00	.00	.02	.00	.01	.

[a]Except ratings.

[b]SK = School knowledge; JK = Job knowledge; AWB = Army-Wide Behaviorally Anchored Rating Scale (BARS).

CONCLUDING COMMENTS

The decision rules and imputation procedures used with the CV data were successful in allowing development of performance scores for a high proportion of the soldiers tested. Based on the available evidence, no significant distortions seem to have been introduced while achieving this goal. Relatively few values were imputed, and where imputation was necessary it was carried out systematically and with a clear rationale.

The apparent ease with which imputation procedures were applied should not, however, lead to relaxation of data collection procedures in the future. Lessons learned from investigation of the reasons for missing data will be used to modify data collection procedures for the Project A Longitudinal Validation, in order to further reduce the amount of missing data.

Chapter 4

DEVELOPMENT OF BASIC PREDICTOR SCORES FOR THE TRIAL BATTERY[1]

Three general principles, consonant with the theoretical and practical orientation that had been employed since the inception of Project A, guided the development of the Predictor Trial Battery:

1. Maximize the heterogeneity of the battery by retaining measures of as many different constructs as possible.

2. Maximize the chances of incremental validity and classification efficiency.

3. Retain measures with adequate reliability.

Taking into account all the information accumulated from development and testing activities, the 6.5-hour Pilot Trial Battery used in the field tests was reduced to the 4-hour Trial Battery for use in Concurrent Validation. Decisions on the final revisions were made in a series of meetings attended by the project staff and the Scientific Advisory Group. Considerable discussion was generated at these meetings, but the group was able to reach a consensus on the reductions and revisions to be made. Table 4.1 shows the array of measures that made up the CV Trial Battery. (See Peterson, 1987, for a complete description of all research activities leading up through the development of the Trial Battery.)

PSYCHOMETRIC PROPERTIES OF THE TRIAL BATTERY

As described in Chapter 2, the Trial Battery was administered to 9,500 MOS incumbents in the Concurrent Validation sample. Test-retest data (2-week interval) were also collected on a subset of about 500 soldiers.

A total of 69 scores were generated from the Trial Battery. Forty-three of these came from the non-cognitive inventories--Assessment of Background and Life Experiences (ABLE), the Army Vocational Interest Career Examination (AVOICE), and the Job Orientation Blank (JOB), which had been included in the AVOICE for the Trial Battery. Six scores came from the six paper-and-pencil, cognitive tests. For the computer-administered tests, a number of alternative methods of scoring, such as slopes, intercepts, and slightly different methods of computing means (priority, different procedures for trimming items before computing means), were evaluated.

---

[1]This chapter is based on excerpts from two papers: (a) Identification of Predictor Constructs and Development of New Selection/Classification Tests, by Norman G. Peterson, Leaetta M. Hough, Marvin D. Dunnette, Rodney L. Rosse, Janis S. Houston, Jody L. Toquam, and Hilda Wing; (b) Project A Validity Results: The Relationship Between Predictor and Criterion Domains, by Jeffrey J. McHenry, Leaetta M. Hough, Jody L. Toquam, Mary Ann Hanson, and Steven Ashworth.

Table 4.1

Description of Measures in the Predictor Trial Battery

COGNITIVE PAPER-AND-PENCIL TESTS	Number of Items	Time Limit (minutes)
Reasoning Test	30	12
Object Rotation Test	90	7.5
Orientation Test	24	10
Maze Test	24	5.5
Map Test	20	12
Assembling Objects Test	32	16

COMPUTER-ADMINISTERED TESTS	Number of Items	Approximate Time
Demographics	2	4
Reacation Time 1	15	2
Reaction Time 2	30	3
Memory Test	36	7
Target Tracking Test 1	18	8
Perceptual Speed and Accuracy Test	36	6
Target Tracking Test 2	18	7
Number Memory Test	28	10
Cannon Shoot Test	36	7
Target Identification Test	36	4
Target Shoot Test	30	5

NON-COGNITIVE PAPER-AND-PENCIL INVENTORIES	Number of Items	Approximate Time
Assessment of Background and Life Experiences (ABLE)	209	35
Army Vocational Interest Career Examination (AVOICE)	176	20

Generally speaking, the scores selected for additional analyses were those that were most reliable and could be interpreted in a straightforward way.

Table 4.2 shows Ns, means, standard deviations, reliabilities, and uniqueness (from ASVAB) coefficients for scores on the cognitive, paper-and-pencil tests. Tables 4.3 and 4.4 show similar data for the computer-administered tests. Tables 4.5, 4.6, and 4.7 show similar data for the ABLE, AVOICE, and JOB scale scores (uniqueness coefficients are not shown for these instruments, but range from .40 to .88, with a median U2 of .79 for ABLE, .80 for AVOICE, and .57 for JOB).

Table 4.2

Concurrent Validity Data Analysis: Statistics for Paper-and-Pencil Cognitive Tests

Test	N	Mean	SD	Split-Half Reliability[a]	Test Retest Reliability[b]	Uniqueness Estimate
Assembling Objects	9,343	23.3	6.71	.91	.70	.65
Object Rotation	9,345	62.4	19.06	.99	.72	.81
Maze	9,344	16.4	4.77	.96	.70	.74
Orientation	9,341	11.0	6.18	.89	.70	.60
Map	9,343	7.7	5.51	.90	.78	.46
Reasoning	9,332	19.1	5.67	.87	.65	.53

[a]Split-half reliability estimates were calculated using the odd-even procedure with the Spearman-Brown correction for test length.

[b]Test-retest reliability estimates are based on a sample of 468 to 487 subjects.

Table 4.3

Concurrent Validity Data Analysis: Statistics for Computerized Psychomotor Tests

Measure	N	Mean[a]	SD	Odd-Even Reliability	Test-Retest Reliability[b]	Uniqueness Estimate
Target Tracking 1						
Mean Log (Distance + 1)	9,251	2.98	.49	.98	.74	.82
Target Tracking 2						
Mean Log (Distance + 1)	9,239	3.70	.51	.98	.85	.79
Target Shoot						
Mean Log (Distance +1)	8,892	2.17	.24	.74	.37	.70
Mean Time to Fire	8,892	235.39	47.78	.85	.58	.78
Cannon Shoot						
Mean Absolute Time Discrepancy	9,234	43.94	9.57	.65	.52	.56

[a]Time-to-fire and time-discrepancy measures are in hundredths of seconds. Logs are natural logs.

[b]Test-retest reliability estimates are based on sample sizes of 468 to 487.

Table 4.4

Concurrent Validity Data Analysis: Statistics for Computerized Cognitive Perceptual Tests

Measure	N	Mean[a]	SD	Odd-Even Reliability	Test-Retest Reliability[b]	Uniqueness Estimate
Simple Reaction Time (SRT)						
Decision Time Mean	9,255	31.84	14.82	.88	.23	.87
Proportion Correct	9,255	.98	.04	.46	.02	.44
Choice Reaction Time (CRT)						
Decision Time Mean	9,269	40.93	9.77	.97	.69	.93
Proportion Correct	9,269	.98	.03	.57	.23	.55
Short-Term Memory (STM)						
Decision Time Mean	9,149	87.72	24.03	.96	.66	.93
Proportion Correct	9,149	.89	.08	.60	.41	.55
Perceptual Speed & Accuracy (PSA)						
Decision Time Mean	9,244	236.91	63.38	.94	.63	.92
Proportion Correct	9,244	.87	.08	.65	.51	.61
Target Identification (TID)						
Decision Time Mean	9,105	193.65	63.13	.97	.78	.83
Proportion Correct	9,105	.91	.07	.62	.40	.59
Number Memory						
Final Response Time Mean	9,099	160.70	42.63	.88	.62	.67
Input Response Time Mean	9,099	142.84	55.24	.95	.47	.85
Operations Response Time Mean[c]	9,099	233.10	79.72	.93	.73	.66
Proportion Correct	9,099	.90	.09	.59	.53	.39
SRT-CRT-STM-PSA-TID						
Pooled Mean Movement Time[c]	8,962	33.61	8.03	.74	.66	.71

[a]Times are given in hundredths of seconds.

[b]N = 460 - 479 for test-retest correlations.

[c]Coefficient Alpha reliability estimates.

Table 4.5

ABLE Scale Statistics for Total Group[a]: Trial Battery

ABLE Scale	No. Items	N	Mean	SD	Median Item-Total Correlation	Internal Consistency Reliability (Alpha)	Test-Retest Reliability[b]
Substantive Scales							
Emotional Stability	17	8,522	39.0	5.45	.39	.81	.74
Self-Esteem	12	8,472	28.4	3.70	.39	.74	.78
Cooperativeness	18	8,494	41.9	5.28	.39	.81	.76
Conscientiousness	15	8,504	35.1	4.31	.34	.72	.74
Nondelinquency	20	8,482	44.2	5.91	.36	.81	.80
Traditional Values	11	8,461	26.6	3.72	.36	.69	.74
Work Orientation	19	8,498	42.9	6.06	.41	.84	.78
Internal Control	16	8,485	38.0	5.11	.39	.78	.69
Energy Level	21	8,488	48.4	5.97	.38	.82	.78
Dominance	12	8,477	27.0	4.28	.44	.80	.79
Physical Condition	6	8,500	14.0	3.04	.60	.84	.85
Response Validity Scales							
Unlikely Virtues	11	8,511	15.5	3.04	.34	.63	.63
Self-Knowledge	11	8,508	25.4	3.33	.36	.65	.64
Non-Random Response	8	8,559	7.7	0.59			.30
Poor Impression	23	8,492	1.5	1.85	.20	.63	.61

[a]Total group after screening for missing data and random responding.

[b]N = 408 - 412 for test-retest correlation (N = 414 for Non-Random Response test-retest correlation).

Table 4.6

AVOICE Scale Statistics for Total Group[a]: Trial Battery

AVOICE Scale	No. Items	N	Mean	SD	Median Item-Total Correlation	Internal Consistency Reliability (Alpha)	Test-Retest Reliability[b]
Clerical/Administrative	14	8,463	39.6	10.81	.67	.92	.78
Mechanics	10	8,382	32.1	9.42	.80	.94	.82
Heavy Construction	13	8,488	39.3	10.54	.68	.92	.84
Electronics	12	8,359	38.4	10.22	.70	.94	.81
Combat	10	8,466	26.5	8.35	.65	.90	.73
Medical Services	12	8,364	36.9	9.54	.68	.92	.78
Rugged Individualism	15	8,396	53.3	11.44	.58	.90	.81
Leadership/Guidance	12	8,444	40.1	8.63	.62	.89	.72
Law Enforcement	8	8,471	24.7	7.37	.65	.89	.84
Food Service - Professional	8	8,472	20.2	6.50	.67	.89	.75
Firearms Enthusiast	7	8,397	23.0	6.36	.66	.89	.80
Science/Chemical	6	8,468	16.9	5.33	.70	.85	.74
Drafting	6	8,493	19.4	4.97	.66	.84	.74
Audiographics	5	8,473	17.6	4.09	.69	.83	.75
Aesthetics	5	8,413	14.2	4.13	.59	.79	.73
Data Processing	4	8,224	14.0	3.99	.78	.90	.77
Food Service - Employee	3	8,304	5.1	2.08	.54	.73	.56
Mathematics	3	8,421	9.6	3.09	.78	.88	.75
Electronic Communications	6	8,403	18.4	4.66	.60	.83	.68
Warehousing/Shipping	2	8,407	5.8	1.75	.44	.61	.54
Fire Protection	2	8,431	6.1	1.96	.62	.76	.67
Vehicle/Equipment Operator	3	8,378	8.8	2.65	.51	.70	.68

[a]Total group after screening for missing data and random responding.

[b]N = 389 - 409 for test-retest correlation.

Table 4.7

JOB Scale Statistics for Total Group[a]: Trial Battery

JOB	No. Items	N	Mean	SD	Median Item-Total Correlation	Internal Consistency Reliability (Alpha)
Job Security	10	7,809	43.6	4.51	.54	.84
Job Pride	5	7,817	21.6	2.33	.43	.67
Serving Others	3	7,784	12.1	1.83	.52	.66
Autonomy	4	7,817	15.1	2.29	.31	.50
Routine	4	7,707	9.6	2.30	.25	.46
Ambition	3	7,751	12.4	1.63	.35	.49

[a]Total group after screening for missing data and random responding.

As these tables show, the battery possesses excellent psychometric properties with the exception of low reliabilities on a few computer-administered test scores. As anticipated, these low reliabilities tend to be characteristic of the proportion of correct scores. That is, the items can almost always be answered correctly if the examinee takes enough time and this situation operates to severely restrict the range on the proportion correct scores. However, it increases the variance (and reliability) on the decision time scores.

At this stage our progress toward the original objectives of predictor measurement in Project A seemed to be as follows:

1. Identify "best bet" measures. This objective has been met. As described in earlier reports, we sifted through all the available literature (Hough, 1986; McHenry & Rose, 1986; Toquam, Corpe, Dunnette, & Keyes, 1986). We then translated the information onto a common form that enabled us to evaluate constructs and measures in terms of several psychometric and pragmatic criteria. The results of that effort fed into the expert judgment process wherein

35 personnel psychologists provided the data necessary to develop our first model of the predictor space. After further review by experienced researchers in the Army and the Scientific Advisory Group, a set of "best bet" constructs was identified. We also made some field visits to observe combat arms jobs first-hand. All of this information contributed to the development of new measures.

2. Develop measures of "best bet" predictors. This objective was accomplished by following the blueprint provided from the first objective. We carried out many tryouts of these measures as they were developed. The Trial Battery is the outcome of meeting this objective.

3. Estimate reliability and vulnerability of measures. Analyses to date indicate that the new measures are psychometrically sound and can be protected from various sources of measurement problems. However, additional research on the effects of response biases, and on ways to account for such effects, is necessary before the tests are used operationally.

4. Determine the interrelationships between the new measures and current pre-enlistment measures. The data collected to date show that a considerable proportion of the reliable variance of the new measures is not shared with the ASVAB, and that the across-domain covariance is low (e.g., the new cognitive measures have low correlations with the non-cognitive measures).

FORMATION OF PREDICTOR COMPOSITES

The preliminary analyses of the new Trial Battery predictor tests indicated that reliable predictor scores could be computed from the six spatial tests (i.e., the paper-and-pencil, cognitive tests), the 10 computerized tests, and the temperament, vocational interest, and job reward inventories (Peterson et al., 1987). In addition, scores from the nine ASVAB subtests were available from Army records. Table 4.8 shows how these predictor scores were distributed among various domains within the predictor space. The ASVAB subtests measured nine cognitive abilities. The paper-and-pencil cognitive tests measured six different aspects of spatial ability. The 10 computerized tests yielded 20 measures of perceptual-psychomotor abilities. The ABLE provided measures of 11 temperament/biographical traits. The AVOICE assessed 22 vocational interests. Finally, the JOB measured six types of job reward preferences.

Several problems precluded using these 78 scores directly in the Project A validity analyses. First, as Table 4.9 shows, the number of subjects with complete predictor and criterion data within the nine target Project A jobs ranged from 289 for Single Channel Radio Operator (MOS 31C)

Table 4.8

Assessment of the Selected Measures With Reference to the Predictor Space

Predictor Domain	Measures[a]	Number of Test or Scale Scores	Number of Composite Scores
General Cognitive Ability	Armed Services Vocational Aptitude Battery (ASVAB)	9 Subtests	4
Spatial Ability	Spatial Test Battery	6 Tests	1
Perceptual-Psychomotor Abilities	Computerized Battery	20 Tests	6
Temperament/Personality	Assessment of Background and Life Experiences (ABLE)	11 Scales[b]	4
Vocational Interests	Army Vocational Interest Career Examination (AVOICE)	22 Scales	6
Job Reward Preferences	Job Orientation Blank (JOB)	6 Scales	3

[a]All measures except the ASVAB were developed specifically for Project A.

[b]The ABLE included four additional response validity scales.

Table 4.9

Number of Subjects With Complete CV Data in the Nine Army Enlisted Jobs Studied

Enlisted Job	MOS	Number of Incumbents
Infantryman	11B	491
Cannon Crewman	13B	464
Armor Crewman	19E	394
Single Channel Radio Operator	31C	289
Light Wheel Vehicle Mechanic	63B	478
Motor Transport Operator	64C	507
Administrative Specialist	71L	427
Medical Specialist	91A	392
Military Police	95B	597

to 597 for Military Police (MOS 95B) (Young, Harris, Hoffman, & Houston, 1987). Even for Military Police, the ratio of subjects to variables was only 8:1. Our intent was to use multiple regression to estimate the correlation between the predictors and job performance constructs. However, the obtained ratio is far less than the ratio of 10:1 that many statisticians regard as the minimum necessary to obtain stable estimates of multiple regression coefficients and the coefficient of multiple correlation $\underline{R}$. Since we were faced with a fixed number of subjects, the only way to improve this ratio was to reduce the number of predictor scores.

Second, scores from any of the predictor tests were highly intercorrelated. For example, the average intercorrelation among the six Project A spatial tests was .46. This multicollinearity results in unstable estimates of multiple regression coefficients. This situation can be remedied by combining the correlated test scores into a single composite. To the extent that the tests are highly intercorrelated, the composite score should contain all of the reliable variance included in any of the individual test scores. Also, the composite should be more reliable than any of the individual test scores, since it will be based on more items than the score from any single test.

Because of these problems, the 78 predictor test and scale scores were combined into 24 predictor composites before predictor-criterion relationships were explored. With one exception (which will be noted below), these composites were formed simply by summing standardized test or scale scores; that is, in all instances but one, unit weights were used to compute composite scores from test and scale scores.

Three principles guided the formation of composite scores. First, we tried to keep the number of composites to a minimum. We expected that this would increase the stability of all of the multivariate statistics we intended to compute in exploring predictor-criterion relationships. Second, we sought to maintain homogeneity of internal consistency within composites. For guidance in this effort, we studied the intercorrelations among test or scale scores. We also used principal components analysis to identify tests or scales with similar patterns of factor loadings. We tended to group test or scale scores with reasonably high intercorrelations and similar patterns of factor loadings into the same composite; we expected that this practice would eliminate any problems associated with predictor multicollinearity. Third, even if we found that two or more test or scale scores were reasonably highly correlated and had similar patterns of factor loadings, we grouped them into the same composite only if we expected that they would have similar patterns of correlations with our job performance constructs. Expert judgments of expected predictor-criterion relationships were available to direct us in this task (Wing, Peterson, & Hoffman, 1984).

Figure 4.1 shows how the nine ASVAB subtests were combined into four composite scores: Technical, Quantitative, Verbal, and Speed. In computing the Technical composite score, the Electronics Information subtest received a weight of one-half unit while the Mechanical Comprehension and Auto Shop subtests received unit weights, because a factor analysis indicated that the loading of the Electronics Information subtest on the Technical factor of the ASVAB was only about one-half as large as the loading of the Mechanical Comprehension and Auto Shop subtests.

As noted above, the six spatial tests were all highly intercorrelated. Therefore, as Figure 4.2 shows, these six tests were combined into a single composite score.

Six composite scores were computed from the 20 perceptual-psychomotor test scores from the computerized battery. These six composites were Psychomotor, Complex Perceptual Speed, Complex Perceptual Accuracy, Number Speed and Accuracy, Simple Reaction Speed, and Simple Reaction Accuracy. Figure 4.3 shows how the 20 test scores were combined into these six composites.

Four temperament composites were computed from the ABLE scales (see Figure 4.4). The composites included Achievement Orientation, Dependability, Adjustment, and Physical Condition. Four of the 11 ABLE scales were not included in any composite.

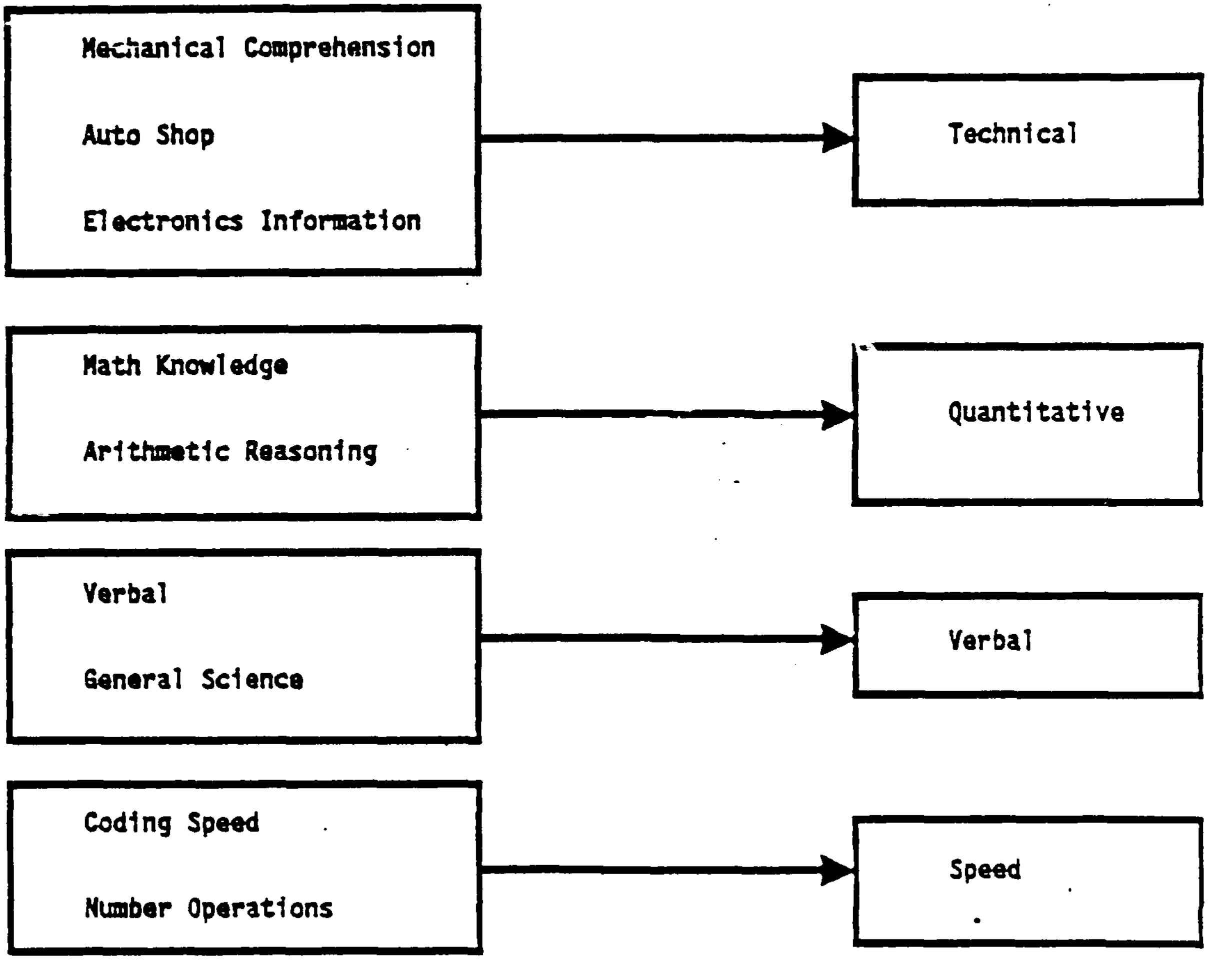


Figure 4.1. Formation of general cognitive ability composites from ASVAB subtests.

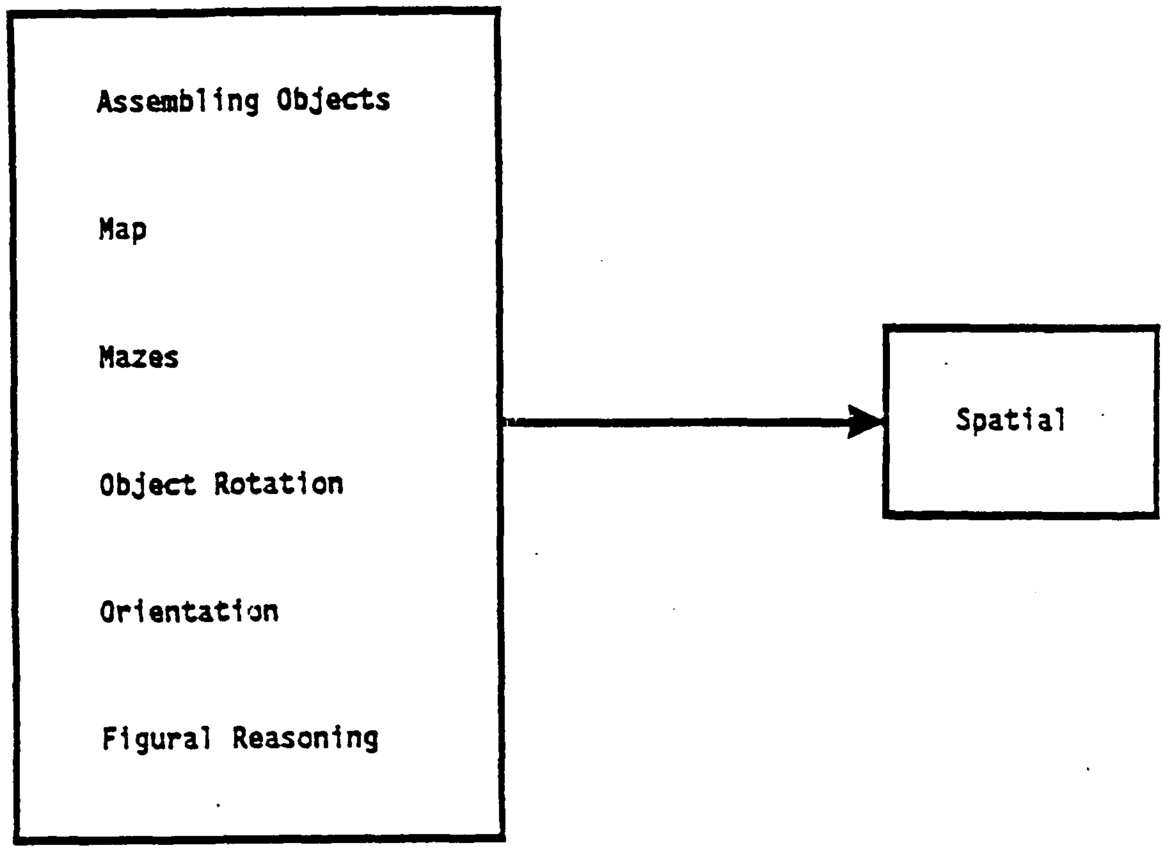
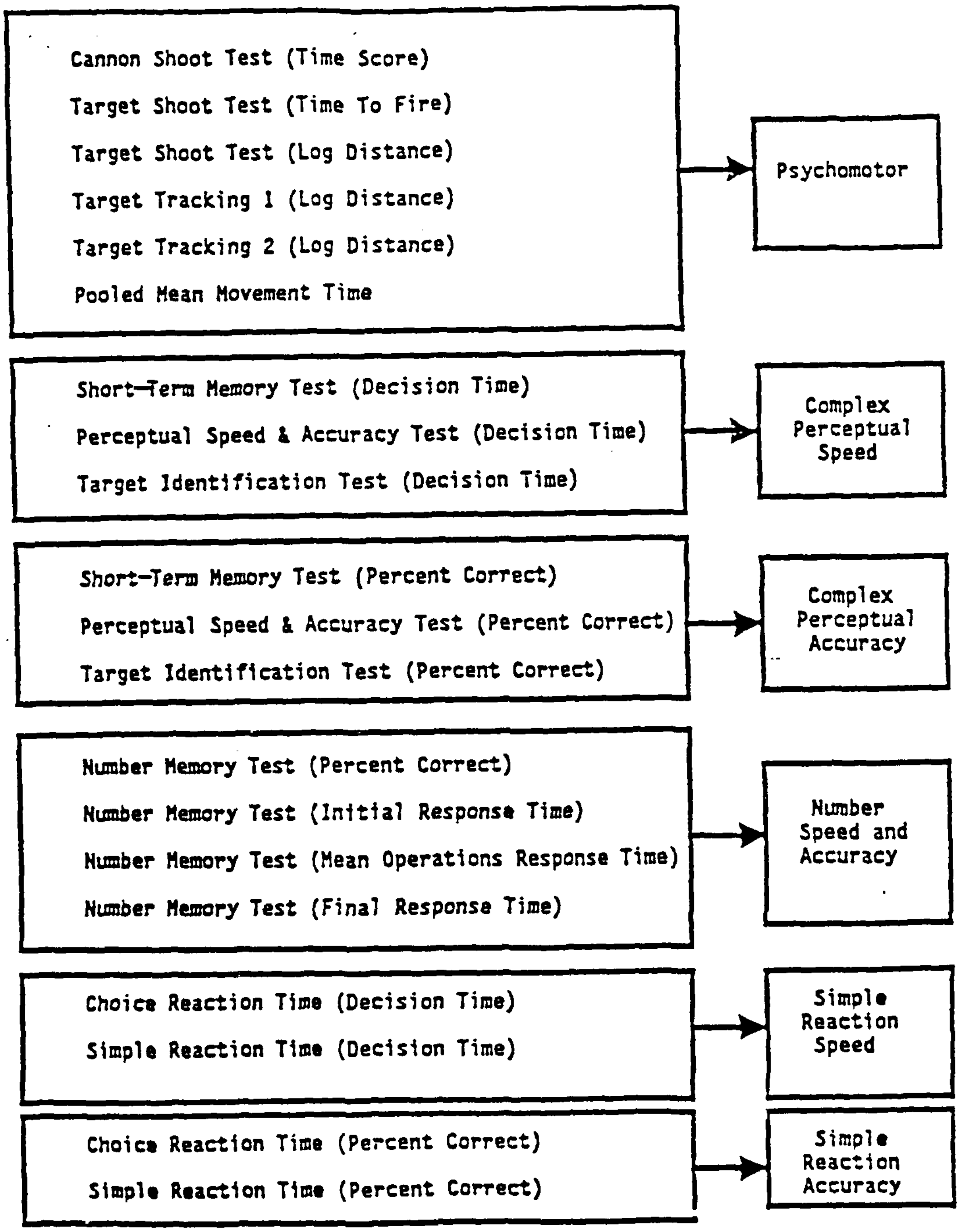
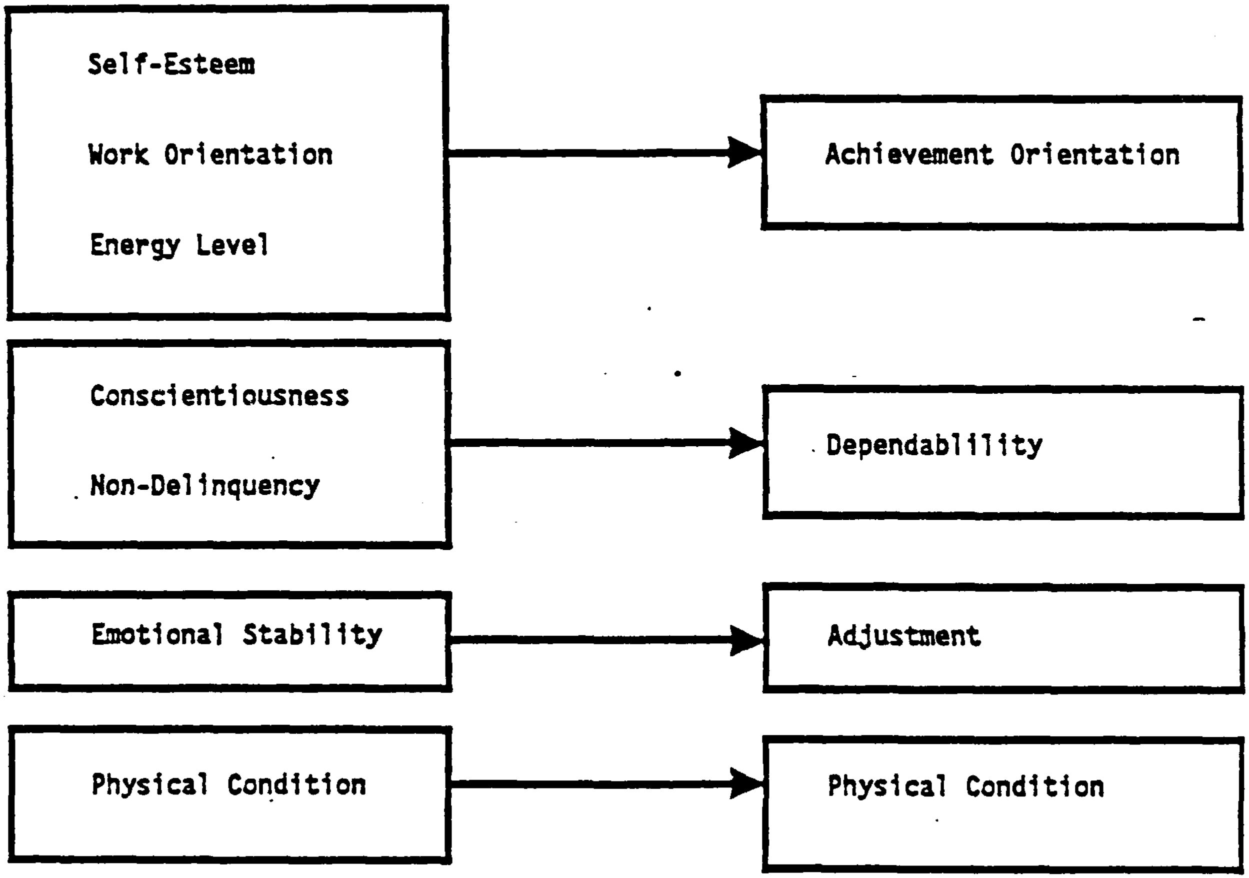


Figure 4.2. Formation of spatial ability composite from spatial battery test scores.



Note: One computer test score, Choice Reaction Time, Decision Time Minus Simple Reaction Time Decision Time was not used in computing composite scores.

Figure 4.3. Formation of perceptual-psychomotor ability composites from computerized battery test scores.



Note: Four ABLE scales (Dominance, Traditional Values, Cooperativeness, and Internal Control) were not used in computing composite scores.

Figure 4.4. Formation of temperament composites from ABLE scale scores.

Figure 4.5 shows that six vocational interest composites were computed from the 21 AVOICE scales. These composites were Skilled Technical, Structural/Machines, Combat-Related, Audiovisual Arts, Food Service, and Protective Services.

Finally, the six scales of the JOB were combined into three composites: Organizational and Co-Worker Support, Routine Work, and Job Autonomy (Figure 4.6).

CONCLUDING COMMENTS

All of our previous predictor score development efforts have produced a description of applicant individual differences in terms of 24 basic scores. These are the scores on which all subsequent predictor validation will be based. They are portrayed in summary form in Table 4.10. The tests and inventory scales from the Trial Battery which were used to form simple sum factor scores are listed under each factor title.

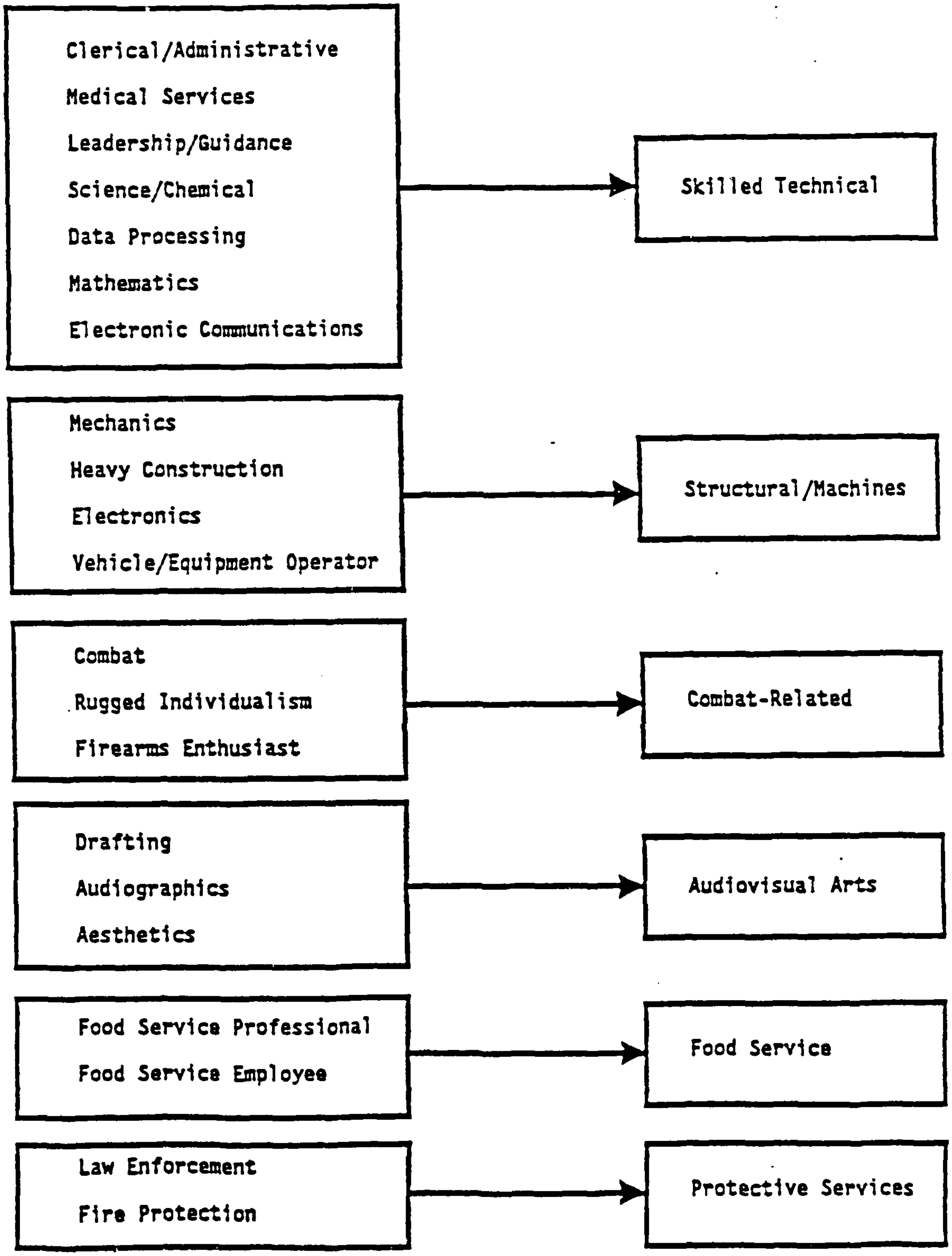


Figure 4.5. Formation of vocational interest composites from AVOICE scale scores.

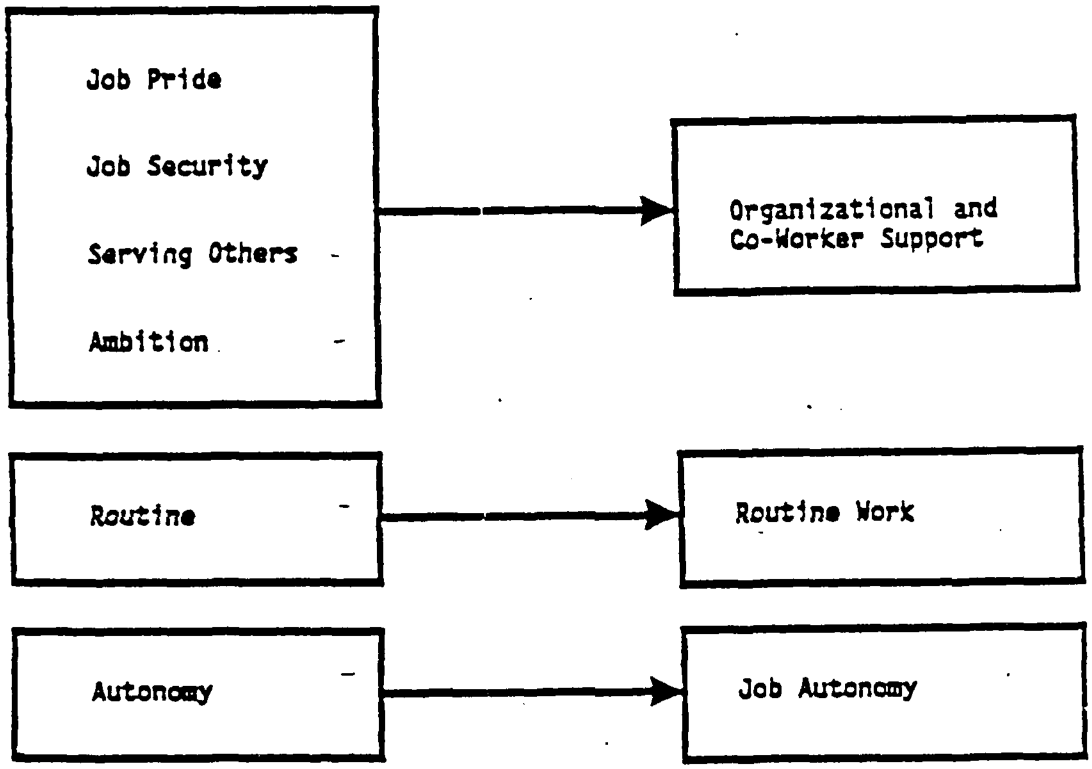


Figure 4.6. Formation of job reward preference composites from JOB scale scores.

Table 4.10

Ability, Temperament, and Interest Factors Identified via Analysis of the Concurrent Validition Data on 9,430 MOS Incumbents

FROM ASVAB SUBTESTS

Technical Factor
- Mechanical Comprehension
- Auto Shop
- Electronics Information

Quantitative Factor
- Math Knowledge
- Arithmetic Reasoning

Verbal Factor
- Verbal
- General Science

Speed Factor
- Coding Speed
- Number Operations

FROM PAPER-AND-PENCIL TESTS

Overall Spatial Factor
- Assembling Objects Test
- Map Test
- Maze Test
- Object Rotation Test
- Orientation Test
- Figural Reasoning Test

FROM COMPUTERIZED MEASURES

Psychomotor Factor
- Cannon Shoot Test (Time score)
- Target Shoot Test (Time to fire)
- Target Shoot Test (Log distance)
- Target Tracking 1 (Log distance)
- Target Tracking 2 (Log distance)
- Pooled Mean Movement Time

Perceptual Speed Factor
- Short-Term Memory Test (Decision time)
- Perceptual Speed & Accuracy Test (Decision time)
- Target Identification Test (Decision time)

Perceptual Accuracy Factor
- Short-Term Memory Test (Percent correct)
- Perceptual Speed & Accuracy Test (Percent correct)
- Target Identification Test (Percent correct)

Number Speed/Accuracy Factor
- Number Memory Test (Percent correct)
- Number Memory Test (Initial decision time)
- Number Memory Test (Mean operations time)
- Number Memory Test (Final decision time)

Simple Reaction Speed Factor
- Choice Reaction Time (Decision time)
- Simple Reaction Time (Decision time)

Simple Reaction Accuracy Factor
- Choice Reaction Time (Percent correct)
- Simple Reaction Time (Percent correct)

FROM NON-COGNITIVE INVENTORIES

Achievement Factor
- Self-Esteem scale
- Work Orientation scale
- Energy Level scale

Dependability Factor
- Conscientiousness scale
- Non-delinquency scale

Adjustment Factor
- Emotional Stability scale

Physical Condition Factor
- Physical Condition scale

Skilled Technical Interest Factor
- Clerical/Administrative
- Medical Services
- Leadership/Guidance
- Science/Chemical
- Data Processing
- Mathematics
- Electronic Communications

Structural/Machines Interest Factor
- Mechanics
- Heavy Construction
- Electronics
- Vehicle/Equipment Operator

Combat-Related Interest Factor
- Combat
- Rugged Individualism
- Firearms Enthusiast

Audiovisual Arts Interest Factor
- Drafting
- Audiographics
- Aesthetics

Food Service Interest Factor
- Food Service Professional
- Food Service Employee

Protective Services Interest Factor
- Law Enforcement
- Fire Protection

Preference for Organizational & Co-worker Support
- Job Pride
- Job Security
- Serving Others
- Ambition

Preference for Routine Work
- Routine

Preference for Job Autonomy
- Autonomy

Chapter 5

DEVELOPMENT OF PERFORMANCE CRITERION SCORES FROM HANDS-ON AND KNOWLEDGE TESTS[1]

This chapter outlines the procedures used to formulate basic criterion scores for the hands-on tests, job knowledge tests, and task rating scales under development in Project A. To that end, we had two specific objectives:

1. To prepare the data for analysis by eliminating extraneous sources of variance and

2. To combine the initial criterion scores into a shorter and more usable list of aggregated criterion scores, for purposes of modeling job performance (see Chapter 7).

DEFINITION OF TERMS

Certain terms used throughout this chapter have meaning that is specific to the content. As an aid to an understanding of the processes involved, the more critical terms are defined below:

Common task: A task drawn from the Soldier's Manual of Common Tasks. These are typically basic soldiering tasks (first aid, personal weapons, map reading, etc.) that are required of all soldiers, regardless of MOS.

Hands-on component: The full set of hands-on performance task tests for an MOS; consists of 14-17 task tests.

Job knowledge component: The full set of job knowledge tests for an MOS; consists of 28-30 task tests.

Rating instrument: The full set of task rating scales for an MOS; consists of 14-17 scales. The tasks included for each MOS are those that are also tested in the hands-on component.

Task: A discrete set of behaviors performed to accomplish a single job requirement; includes a situation with initiating cues and conditions, the steps or activities that are to be performed, and the task standards that signal successful completion.

---

[1]The materials in this chapter were drawn from Developing Basic Criterion Scores for Hands-On Tests, Job Knowledge Tests, and Task Ratings Scales, by Charlotte H. Campbell. (HumRRO IR-PRD-87-15) (ARI Technical Report in preparation).

Task rating scale: A 7-point scale that deals with a single task. The scale anchors are 1 = "Among the very worst," to 7 = "Among the very best." A response point of 0 = "Not observed" is also available. Rating scales are filled out for each soldier by his or her peers and supervisors.

Task test: A set of hands-on performance steps or written job knowledge items that are focused on a single task. Hands-on steps (from 4 to 62 per task) are scored GO or NO-GO by a trained scorer. Job knowledge test items (from 2 to 16 per task) are multiple choice, with one correct answer.

Technical task: A task that is central to the job of soldiers in an MOS, and is typically unique to the MOS.

Test mode: Either hands-on or knowledge.

Track: A separate version of a hands-on or knowledge test prepared to accommodate different types of equipment that may be used to perform the same task.

MOS TASK-SPECIFIC CRITERION CONTENT

The content of each criterion measure and how it was developed are described in detail in the Project A Annual Report for 1985 (Campbell, 1987). However, for reference purposes, the task contents of the hands-on tests and the job knowledge tests for the nine Batch A MOS are shown in Appendix B.

As described in earlier reports, test content was generated by using all available information to define a population of tasks for each MOS, obtaining judgments by subject matter experts (SME) on several task parameters, systematically sampling from the task population, and submitting task samples to multiple reviews by the proponent. Multiple-choice job knowledge test items were generated for 30 tasks per MOS, and hands-on test stations were developed to test performance on 13-16 of these tasks per MOS. Supervisor and peer ratings were also obtained in an attempt to assess typical proficiency on the tasks that were also measured hands-on; that is, how well did the individual typically perform a particular task back on the job?

ADJUSTMENT OF DATA

There were three known sources of variance in the data which we considered extraneous: tracking, site differences, and test length. Procedures were developed to adjust data to minimize the effects of these sources of variance.

Hands-on tracked tests were prepared in one of three forms, depending on how much the tracks differed. In some cases, where equipment variations required only minor differences in the performance of a few hands-on steps, or the omission of a few steps, the separate tracks were covered in a single version of the hands-on test. Where alternate steps were listed, the soldier was scored on only those steps appropriate to his or her track; where a few steps had to be omitted for one track, those steps were imputed. (See Chapter 3.)

In other cases, the procedures for performance differed for the tracks but the number of scorable steps for the tracks was the same, and the behavioral requirements and equipment configurations were very similar. In these cases the decision was made to prepare the separate versions but to consider them as equivalent; for each track, the soldier's percentage GO, or number of steps GO, score was used without further equating.

In still other cases, equipment and procedures were so dissimilar that we were not convinced that the behavioral requirements were the same across the tracks, even though the tracks represented the jobs equally well. It was apparent that there were likely to be level and dispersion differences in the test scores that would reflect task difficulties rather than individual differences among soldiers. For these tracked tasks, the scores were standardized by reference to the other technical (as opposed to common) tasks in the hands-on component for the MOS.

For the job knowledge tests, fewer tracked tests were needed because in most cases generic items could be prepared, without reference to specific equipment models. Where tracked tests were developed, the scores on the tracks were standardized by reference to the other technical tasks in the job knowledge component for the MOS.

In the hands-on tests, sources of variance included not only the individual differences among soldiers but also the test site, the soldier's unit (nested within test site), and the test scorer (also nested within site). Initial analyses (see Hoffman, 1986) indicated that adjusting hands-on scores for site differences, by standardizing across test sites, would eliminate anomalies in the hands-on data.

A third reason for adjusting the hands-on scores was that hands-on task tests varied widely in the number of scorable steps. Because we did not feel that this aspect reflected the difficulty, complexity, or criticality of the task, the step data were adjusted by weighting them as though each task had 14 steps (14 being the average number of steps per task across the nine MOS). The adjusted step data were then used in analyses.

Job knowledge tests were much less variable, so no similar adjustment to items was made.

STATISTICAL CHARACTERISTICS OF MEASURES

Table 5.1 shows the number of hands-on task tests and steps, the number of job knowledge task tests and items, the number of rating scales, and the range of raters per ratee among the scales for each of the nine MOS. These are presented to support the statement of statistical characteristics of the tests shown in Table 5.2 for each of the three measures for the nine MOS.

Table 5.1

Number of Job Knowledge Tests and Items, Hands-On Tests and Steps, and Task Rating Scales and Raters-Per-Ratee for Nine MOS

Measure	MOS 11B	13B	19E	31C	63B	64C	71L	91A	95B
Job Knowledge									
Task Tests	28	30	29	30	30	28	25	30	31
Items[a]	176-187	178-182	168-192	206	194	167	144	230	210-213
Hands-On									
Task Tests	13-14	17	15	15	15	16	15	15	16
Steps[a]	209-241	308-315	184-206	353	172	284	187	220	276-305
Steps (Adjusted)[b]	182-196	238	210	210	210	224	210	210	224
Rating Scales									
Scales	14	17	15	15	15	16	15	15	16
Rater/Ratee Range[c]	2.28-4.61	2.34-4.14	2.35-4.24	2.20-3.49	2.19-3.70	2.27-4.00	1.51-2.47	2.03-3.49	1.99-4.62

[a]The number of items or steps is shown as a range when separate tracks were prepared.

[b]Task step data were weighted to give each test equal weight of 14 steps per test; thus the adjusted number of steps for each MOS is the number of task tests multiplied by 14.

[c]The number of raters per soldier ratee includes both peer and supervisor raters.

Table 5.2

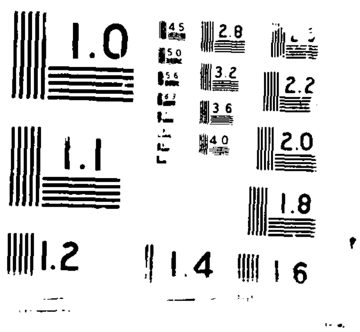
Statistical Characteristics of Hands-On Component, Job Knowledge Component, and Rating Instruments for Nine MOS

		MOS								
Measure		11B	13B	19E	31C	63B	64C	71L	91A	95B
Job Knowledge Task Tests										
Range of Test Means	%	37.4-83.2	34.4-85.2	27.0-88.0	30.0-85.0	38.1-86.7	26.8-80.1	23.1 73.4	39.7-89.1	24.5-81.9
Range of Test SDs	%	16.8-35.6	15.7-38.5	15.3-27.2	15.6-33.6	14.9-41.5	15.1-32.9	17.0-39.1	16.0-30.8	15.4-44.0
Split Half - Tests	r	.89	.85	.89	.86	.87	.85	.82	.89	.84
Component Mean	%	59.9	60.5	64.0	59.9	64.5	58.2	57.5	65.7	61.3
Component SD	%	11.4	10.6	9.4	10.3	10.8	9.9	10.2	10.0	8.7
Alpha		.92	.91	.89	.91	.92	.89	.92	.92	.89
Hands-On Task Tests										
Range of Test Means	%	41.3-94.5	23.8-90.8	49.6-96.7	50.3-85.2	70.8-94.7	37.1-89.3	26.6-83.9	52.1-88.0	20.2-84.5
Range of Test SDs	%	9.7-33.3	14.4-37.5	5.2-33.4	10.0-32.0	12.0-25.9	14.8-33.1	16.5-43.1	11.9-36.5	10.3-26.9
Split Half - Tests	r	.54	.75	.63	.79	.52	.64	.73	.60	.58
Component Mean	%	56.2	61.5	75.4	70.1	84.5	71.5	58.9	71.4	70.2
Component SD	%	10.8	11.0	8.3	7.9	4.8	7.,6	8.5	7.5	6.3
Alpha		.85	.94	.85	.91	.75	.88	.87	.89	.82
Rating Scales[a]										
Range of Scale Means		4.37-5.14	3.94-5.28	4.22-5.37	4.41-5.23	4.43-5.29	4.24-4.99	4.51-5.45	4.59-5.48	3.75-5.32
Range of Scale SDs		.67-1.13	.63-1.14	.64-1.08	.73-1.08	.70-1.00	.58-1.01	.86-1.18	.70-1.00	.63-1.05
Instrument Mean		4.76	4.79	4.88	4.12	4.82	4.64	5.02	5.04	4.82
Instrument SD		.57	.56	.53	.51	.58	.49	.60	.58	.53
Median Scale Reliability (Intraclass Correlation)		.38	.28	.33	.37	.36	.21	.29	.24	.37

[a]Combined supervisor and peer.

AD-A190 056 IMPROVING THE SELECTION CLASSIFICATION AND UTILIZATION 2/3
OF ARMY ENLISTED P..(U) HUMAN RESOURCES RESEARCH
ORGANIZATION ALEXANDRIA VA J P CAMPBELL MAY 88

UNCLASSIFIED HUMRRO-IR-PRD-87-10 ARI-TR-792 F/G 5/9 NL


1.0
4.5
5.0
5.6
2.8
3.2
3.6
4.0
2.2
2.0
1.1
1.8
1.2
1.4
1.6

For the hands-on component, statistics presented include the range of the task test means and standard deviations, where task test scores consist of the percentage of task steps scored GO, and task step scores have been adjusted for task test length. The component mean and standard deviation are based on all steps, across the hands-on tests in the component, again expressed as percentage of steps scored GO with steps adjusted for test length. Two indexes of reliability are given: the corrected split-half estimate using odd vs. even task test scores, and the component alpha computed for each task test using step scores and pooled across tasks. In this situation, coefficient alpha is biased upward since step scores are not independent within tasks; however, they do give some indication that the individual task tests are relatively homogeneous.

For the job knowledge component, the results are presented in a similar format. The individual task test means and standard deviations are expressed as percentage of items passed. The corrected split-half reliability estimate is based on task test scores and the coefficient alpha on item scores.

The task rating scale results include the range of scale means and standard deviations, and the median of the scale reliabilities (intraclass correlations) before ratings were imputed. The mean and standard deviation across scales were computed after imputing.

For job knowledge and hands-on components, the level of difficulty and estimates of reliability seem satisfactory, and there was no evidence of extreme skew in the data. The task rating instrument was far less satisfying; the scale reliabilities were low, as was the average number of raters per ratee.

CONSTRUCTION OF BASIC CRITERION SCORES

As we began development of a basic criterion scoring system for these measures, there was general agreement that one score for each item, each step, and each scale was too many, and that one score for each hands-on task test (15), each job knowledge task test (30), each rating scale (30), and each rater was likewise too many. Conversely, using only a written test total score, a hands-on total score, and an average rating would mask whatever differential task performance existed.

We explored two rational approaches in trying to reduce the number of criterion scores derived from the hands-on tests, job knowledge tests, and task rating scores. One approach concentrated on the functional characteristics of the tasks, the other on the behavioral requirements.

Functional Categories

The task domains for each of the nine MOS, as defined by the Army Occupational Survey Program (AOSP) task lists, were reviewed by the project staff and tasks were clustered into a set of functional categories on the basis of task content. Ten of the categories applied to all MOS and consisted primarily of common tasks. In addition, each MOS, except for 11B (Infantryman) and 64C (Motor Transport Operator), had two to five

MOS-specific categories. The ten common categories were sufficient to account for all tasks in 11B and 64C.

After category definitions had been written, three members of the project staff independently classified the 30 tasks in each MOS into one of the ten common categories or into an MOS-specific category. For each MOS, the three judges included the staff member who had primary responsibility for developing the tests, and one person from each of the two offices where tests were developed. All judges had been involved in test development. The level of perfect agreement in the assignment of tasks to categories was over 90 percent in every MOS. Disagreements were resolved by discussion and subsequent revision of the functional category definitions.

These same functional categories were used by the project staff responsible for the school knowledge tests to sort items into content categories (Kuhn, Schultz, & Park, in preparation). After the training performance test items had been categorized, discussions among project staff led to further revisions in the category definitions. The resulting set of categories was acceptable to both project teams as a way of describing the functional characteristics of an MOS.

The functional category definitions are presented in Figure 5.1 and the MOS task assignments to functional categories are shown in Table 5.3.

Judged Estimates of General Behavioral Requirements

After reviewing the literature on taxonomies for behavioral requirements and job analyses, the project staff developed a set of four knowledge categories and two performance categories that we felt would account for the critical differences in response capabilities required by the hands-on steps and job knowledge items. These definitions are shown in Figure 5.2.

Three project staff then independently sorted each of the hands-on steps and job knowledge test items into one of the six categories. As with the assignments to functional categories, for each MOS the three judges included the staff member who had primary responsibility for developing the tests, and one person from each of the two offices where tests were developed. All judges had been involved in test development. The staff who sorted the hands-on test items all had served as hands-on test managers during the Concurrent Validation testing. The frequency of complete agreement in category assignments among the judges averaged about 80 percent for hands-on and job knowledge tests across the nine MOS. Disagreements were resolved via discussion among the three judges.

The distribution of hands-on tests and knowledge test items across the six knowledge/performance categories is shown in Table 5.4.

Common Categories

First Aid

Consists of items whose primary purpose is to indicate knowledge about how to sustain life, prevent health complications caused by trauma or environmentally induced illness, including the practice of personal hygiene. Includes all related diagnostic, transportation, and treatment items except those items normally performed in a patient care facility. Includes items related to safety and safety hazards.

NBC

Consists of items whose primary purpose is to indicate knowledge about performance when nuclear, biological, or chemical contaminants and threats are present, planned, detected, or expected. Includes maintenance and operation of clothing, gear, and equipment whose primary purpose is to counter, protect, or detect NBC threats. Includes NBC markers. Does not include first-aid treatment of contamination.

Weapons

Consists of items whose primary purpose is to indicate knowledge about maintenance, preparation, and firing of small arms. Small arms are defined as sized weapons, including automatic weapons, up to and including caliber .60 and shotguns. Includes ancillary sighting systems and techniques, stands and mounts, zeroing, and techniques of fire. Excludes firing from aircraft and vehicles where the weapon is fired by electrical/hydraulic aiming/firing systems and sighting systems that are part of the aircraft/vehicle and not part of the weapon.

Navigate

Consists of items whose primary purpose is to indicate knowledge about how to plan or execute movement between points over unknown terrain either cross-country or using road networks, or identify the location of objects. Includes all means of determining direction, distances, and locations using maps of all types, overlays, compasses, terrain, celestial objects, and field expedients.

Field Techniques

Consists of items whose primary purpose is to indicate knowledge about operation orders, battlefield survival in defensive and offensive situations. Includes preparation of fighting positions, individual

Figure 5.1. Functional definitions for task categories.
(Page 1 of 7 pages)

concealment, and cover. Includes surveillance, observation, and dismounted movement techniques in all terrain and under hostile conditions. Includes gathering combat intelligence and the practice of counterintelligence. Includes emplacement and detection of mines and boobytraps and employment of hand grenades. Includes how to prepare and maintain ambush sites and move either by vehicle or foot tactically. Includes all items having to do with knowledge about ambushes, and tactical movement on foot. Includes how to enter tactically a building and rooms in a building. Includes all items having to do with knowledge about entering buildings in an urban environment. Includes ways to camouflage or conceal self and motor vehicles from enemy observation in different types of areas (e.g., forests, deserts). Also covers concealing tracks and other distinguishing evidence indicating the presence of a motor vehicle.

Customs and Laws

Consists of items whose primary purpose is to indicate knowledge about the Geneva convention, military SOP governing the treatment of enemy personnel, engagement of the enemy, the conduct of military protocol and ceremony, guard duty, and physical readiness.

Communications

Consists of items whose primary purpose is to indicate knowledge about the facilitation of voice transmissions over tactical wire and FM radios. It includes the use of CEOI and speech security and the maintenance and installation of communication equipment.

Identify Targets

Consists of those items whose primary purpose is to indicate knowledge about vehicle/aircraft recognition.

Anti-Air/Tank Weapons

Consists of items whose primary purpose is to indicate knowledge about how to neutralize enemy tank and air threats.

Vehicle Operation

Consists of items whose primary purpose is to indicate knowledge required to operate vehicles and trailers in usual and unusual conditions, on and off road, alone and in convoy. Items also focus on the proper procedures (e.g., loading, driving) to follow when transporting passengers, weapons, ammunitions, and hazardous cargo. It also covers understanding road signs and hand signals. Includes supervision of loading/unloading and cargo security. Includes reaction to emergency driving conditions. Includes operator maintenance of non-MOS specific wheel or track vehicles, or both. Includes completion of appropriate forms.

Figure 5.1. Functional definitions for task categories.
(Page 2 of 7 pages)

MOS-Specific Categories

13B - Cannon Crewman

Prepare, Operate, Maintain Howitzer and Ammunition

Consists of items whose primary purpose is to indicate knowledge about tactically employing the howitzer and ammunition and maintaining both to insure reliability. Includes preparation of the howitzer position, emplacement of the howitzer and ammunition, driving, and preparing and storing ammunition at the position. Includes operator and crew maintenance, PMCS and disassembly, cleaning, and assembly of howitzer components, and inspection and recording of equipment faults on appropriate forms.

Operate Howitzer Sights and Alignment Devices

Consists of items whose primary purpose is to indicate knowledge required to orient, lay a howiter on an azimuth of fire using the howitzer sights, position aiming posts and the collimator, boresight, and engage direct fire targets from the assistant gunner's position.

19E - Cannon Crewman

Operate Tanks

Consists of items whose primary purpose is to indicate knowledge about tank driving, including items involving automotive and suspension maintenance. Includes the function, operation, and maintenance of tank accessories, such as gas particulate filter system and driver's periscope. Does not include fire control system.

Tank Gunnery

Consists of items whose primary purpose is to indicate knowledge about activities of the gunner, loader, or tank commander in preparation for engagement, conduct of fire, and actions after engagement with the tank weapons system. Includes ammunition and maintenance of the fire control system and reaction to adverse conditions. Does not include maintenance of machineguns but includes installation and techniques of fire of these weapons.

Figure 5.1. Functional definitions for task categories.
(Page 3 of 7 pages)

31C - Single Channel Radio Operator

Generators

Consists of items whose primary purpose is to indicate knowledge associated with the installation, operation, and maintenance of non-automotive generators.

TTY Station and Net Operations

Consists of items whose primary purpose is to indicate knowledge associated with operating within and controlling TTY networks. Encompasses codes, ciphers, correct message construction, handling unclassified and classified messages, station site selection and inspection, personnel assignment, the control of equipment and supplies, safety and security procedures, the preparation and maintenance of logs, records, and files, and the preparation of reports.

Maintain TTY Electronic Equipment

Consists of items whose primary purpose is to indicate knowledge relating to the inspection of radioteletype electronic equipment as part of routine periodic maintenance procedures, the testing of new or repaired equipment; troubleshooting specific problems during operation to identify and repair minor defects or report major defects to maintenance personnel.

Operate TTY Electronic Equipment

Consists of items whose primary purpose is to indicate knowledge about the operation of specific radioteletype equipment including receivers, transmitters, modems, terminals, reperforators and ancillary equipment, including remote control devices, for the purpose of carrying out communication operations.

Install TTY Electronic Equipment

Consists of items whose primary purpose is to indicate knowledge about the assembly, preparation, grounding, and installation of TTY electronic equipment, including the connection of cables, the construction or erection of expedient and non-expedient antennas, and the installation of other ancillary components prior to operation.

Figure 5.1. Functional definitions for task categories.
(Page 4 of 7 pages)

63B - Light Wheel Vehicle Mechanic

Electrical System

Consists of items whose primary purpose is to indicate knowledge about replacement, repair, testing, and troubleshooting of vehicle electrical components. The types of vehicles range from the 1/4-ton to the 5 ton. Equipment worked on includes starter, ignition, lights and horn.

Fuel/Cooling/Lubricating

Consists of items whose primary purpose is to indicate knowledge about vehicle fuel and cooling systems. The vehicles worked on include 1/4-ton, 1-1/4-ton, 2-1/2-ton, and 5-ton. Includes replacing, repairing, and troubleshooting the fuel pump, thermostat, radiator, and oil filter.

Brake/Steering/Suspension Systems

Consists of items whose primary purpose is to indicate knowledge about troubleshooting, adjusting, repairing, and replacing components of the braking, steering, or suspension system of wheeled vehicles up through 5-ton. Includes repair/replacement of brake master cylinder, brake shoes, hand brake, torque rods, shock absorbers, and axle shafts.

Vehicle Operation and Recovery

Consists of items whose primary purpose is to indicate knowledge about the operation and maintenance of organic vehicles including wreckers and shop vans and the operation and maintenance of general and specific tools and equipment assigned to those vehicles. Includes wheeled vehicle recovery and emergency repairs performed in conjunction with field recovery.

71L - Administrative Specialist

Forms/Files Management

Consists of items whose primary purpose is to indicate knowledge about how to prepare and maintain files, forms, publications and correspondence. Includes establishing and updating functional files, reviewing personnel forms, preparing file plans, and requisitioning publications.

Figure 5.1. Functional definitions for task categories.
(Page 5 of 7 pages)

Supervision/Coordination

Consists of items whose primary purpose is to indicate knowledge about the management of personnel, supplies, and office equipment. Includes briefing personnel, recommending awards and discharges, writing reports, counseling personnel, and selecting details.

Correspondence

Consists of items whose primary purpose is to indicate knowledge about preparing and managing military forms, letters, and other official correspondence. Includes typing memos, letters, and disposition forms, dispatching distribution, and assembling correspondence. Also includes answering inquiries and carrying out normal office routine.

Classified Material

Consists of items whose primary purpose is to indicate knowledge about receipt, control, and dispatch of classified documents. Includes safeguarding FOUO material, maintaining an office security plan, inventorying classified documents, and controlling sensitive forms.

91A-Medical Specialist

Clinic/Ward Treatment and Care

Consists of items whose primary purpose is to indicate knowledge about routine and emergency services administered to patients and casualties in a hospital or clinic in either garrison or field settings. It includes day-to-day care, such as reading vital signs, conducting tests or measurements, and administering injections and other medications requested by physician; assisting patients in regaining strength and movement, and in personal care; and responding to real or potential emergencies, such as seizures, poison, tracheal suction.

Clinic/Ward Housekeeping

Consists of items whose primary purpose is to indicate knowledge about how to maintain safe and sanitary conditions in clinic or ward. Includes cleaning of equipment, maintaining supplies, checking safety equipment, temperature, and ventilation.

Clinic/Ward Management

Consists of items whose primary purpose is to indicate knowledge about how to assure patient flow through facility, and the maintenance and processing of patient and clinic records and forms. Includes emergency evacuation of patients. Patient contact involves admitting and discharge paperwork, escorting patients, or briefing them on facility rules and services.

Figure 5.1. Functional definitions for task categories.
(Page 6 of 7 pages)

95B- Military Police

Responding to Alarms

Consists of items whose primary purpose is to indicate knowledge about how to report to the scene of a real or impending crisis and act to control or reduce danger. Includes bomb threats, traffic accidents, domestic disturbances, or other alarms.

Patrol Duties

Consists of items whose primary purpose is to indicate knowledge associated with day-to-day police activities, including traffic control, investigations, searching suspects and buildings, and completing incident reports. Includes procedures on how to control an area in different situations. Situations include: route reconnaissance, circulation control points, establishing and operating a roadblock.

Conduct MP Procedures

Consists of items whose primary purpose is to indicate knowledge about the correct procedures in various MP situations: apprehending suspects, conducting building, vehicle, and person searches, collecting and processing evidence, transporting offenders, reading the Miranda Rights (Article 31), and patrolling. Also includes preparing and filing the MP reports. Includes the guidelines pertinent to how an MP should conduct himself/herself, or to determine proper procedure in handling various MP situations. These guidelines pertain to: force in apprehension and arrest and when to use it, methods and rules of searches, interviewing offenders/witnesses/complainants, testifying in court, and duties and proper conduct of an MP.

Figure 5.1. Functional definitions for task categories.
(Page 7 of 7 pages)

Table 5.3

Number of Tasks Assigned to Functional Categories for Nine MOS

	MOS								
Category	11B	13B	19E	31C	63B	64C	71L	91A	95B
First Aid	3	3	3	4	2	4	2	5	2
NBC	2	3	3	3	2	5	3	3	3
Weapons	7	3	3	2	2	3	2	1	4
Navigate	3	2	2	1	2	2	2	1	2
Field Techniques	13	2	3	1	1	2	2	2	4
Customs & Laws		1	1	1	1	1	1		
Communications	1	1	3	2					2
Identify Targets	1	1	1	1		1		1	
Vehicle Operation				1		12		1	1
Anti-Air/Tank Weapons	2								
Prepare Howitzer		9							
Operate Howitzer Sights		5							
Operate Tanks			5						
Tank Gunnery			6						
Generators				2					
TTY Station Operations				4					
Maintain TTY Equipment				3					
Operate TTY Equipment				3					
Install TTY Equipment				4					
Electrical System					2				
Power Train/Clutch					2				
Fuel/Cooling/Lub.					4				
Brake/Steering/Susp.					6				
Vehicle Op./Recovery					5				
Forms/Files Mng.							3		
Sup./Coordination							2		
Correspondence							9		
Classified Material							2		
Treatment/Care								14	
Clinic Housekeeping								1	
Clinic Management								1	
Respond to Alarms									3
Conduct MP Proc.									2
Patrol Duties									5

Knowledge Categories

K1 - Recognition and Recall of Facts

Entails the recognition of objects and words, recall of specific facts and principles. Memory is the main underlying component of this knowledge category. Cues for K1 items are often words like "How many...", "How often...", "When do you...". Memory is needed.

K2 - Recall of Procedures

Entails recalling a multistep procedure or a step of a procedure. The procedure consists of a well-defined sequence of behaviors where one step follows another. Often (not always) the cue in a test item is in the form of "What should you do next?" or "Before you do X you should...", or "after X, you do...". Memory is needed here too, only more of it.

K3 - Interpretation and Application

Involves using textual or graphic information, interpreting it, and applying it to select a course of action (or answer a test question). Examples: using a Training Manual or Supplement Book to answer questions: reading a troubleshooting chart; reading/interpreting any scale, such as a compass, ruler, meter, dial, or map. Also includes comparing a product with external criterion or a model, and filling in blanks on a printed form.

K4 - Inference From Principle

Refers to the process of inferring from a general principle. It implies the ability to answer a test question (or at least eliminate some obvious wrong choices) without needing to know specific facts. It includes applying a principle to answer a question rather than recall of a specific fact or procedure. Specific guidelines are not provided for every possible application (e.g., how to camouflage a .45 cal pistol, M16, M60, 1-1/4 ton).

Performance Categories

P1 - Simple Motor

Consists of "simple" motor performance where proficiency can be attained in one or two learning trials. The manipulation is trivial, easy to perform, and easily learned (e.g., push button, switch knobs).

P2 - Complex Motor

Requires motor performance further along the complexity continuum that requires practice to perform well. Two types of steps could be classified P2. The most obvious is a step that requires more than two learning trials. The second type is a group of simple steps that must be practiced several times to meet tight time requirements or to perform under constrained conditions, such as cramped quarters.

Figure 5.2. Definition of knowledge and performance Behavioral Requirement categories.

Table 5.4

Number of Job Knowledge Items and Hands-On Steps Assigned to Behavioral Requirements Categories for Nine MOS

Categories	MOS 11B	13B	19E	31C	63B	64C	71L	91A	95B
Job Knowledge Test Items									
Knowledge Categories									
K1 Recognition/Recall Facts	96	82-89	84-99	120	111	75	82	152	104
K2 Recall Procedures	27-35	35-39	35-38	37	46	32	25	41	30-33
K3 Interpretation/ Application	48	49-52	44-50	44	29	44	36	29	63
K4 Inference From Principle	8	7-8	5	5	8	16	1	8	13
Hands-On Test Steps									
Knowledge Categories									
K1 Recognition/Recall Facts	27-29	35-37	18	68	23	39	22	37	32
K2 Recall Procedures	20-26	30-31	16	12	17	21	10	22	46-50
K3 Interpretation/ Application	24	7-8	19-33	157	23	13	25	36-37	71-72
K4 Inference From Principle	6	1	1	2	2	2	1	8	4
Proficiency Categories									
P1 Simple Motor	114-135	152-169	107-114	87	54	140	39	76-78	83-107
P2 Complex Motor	18-21	79-83	23-24	27	53	169	90	39-40	39

Note: When a range of steps/items is shown, it represents the assignments of items or steps to categories from the several tracks in the individual MOS.

SUMMARY OF RESULTS

Task Rating Scales

It became apparent from the initial analyses that the task rating scales were not providing useful information. We had hoped that they would give us another measure of task proficiency, in addition to the job knowledge and hands-on tests. However, they tended to be uncorrelated with knowledge and hands-on measures of the same tasks, and more highly correlated with the Army-wide and MOS-specific rating scales (Pulakos & Borman, 1986b). They also exhibited low reliability and high rates of missing data, and yielded uninterpretable factor analyses. Consequently, the task rating scales were dropped from further analyses.

Functional Categories

Scores were computed for the functional categories by taking the sum of the hands-on task test steps (adjusted for length) or job knowledge test items in each category. The mean category scores, standard deviations, reliability estimates (alpha), and intercorrelations among categories for hands-on and job knowledge components are shown in Appendix C for the nine MOS.

Separate principal components analyses were then carried out for each MOS, using the functional category score intercorrelation matrix as the input. The results of the factor analyses in each of the nine MOS suggested a similar set of category clusters, with minor differences across all nine MOS. The ten functional categories that cut across MOS and the several technical functional categories which were unique to particular MOS were reduced to six clusters:

1. Communications - including the Communications functional category.

2. Vehicles - including the Vehicle Operation functional category, and for MOS 63B only the Vehicle Operation and Recovery category; for MOS 64C, however, the Vehicle Operation functional category went into the Technical cluster.

3. Basic Soldiering - including the Navigate, Weapons, Field Techniques, Customs and Laws, and Anti-Air/Tank Weapons categories.

4. Identify Targets - including the Identify Targets functional category.

5. Safety/Survival - including the First Aid and NBC functional categories.

6. Technical - including the functional categories peculiar to each MOS, comprising (usually) MOS-specific tasks; for MOS 64C, this cluster included the Vehicle Operation category, which comprises tasks central to the 64C job.

Although this set of clusters was not reproduced precisely for every one of the MOS, it appeared to be a reasonable portrayal of the nine jobs when a common set of clusters was imposed on all. Tables 5.5 and 5.6 show the range correlations among the clusters and between the categories and the clusters, across the nine MOS.

Behavioral Requirements

The behavioral requirements categories did not yield any useful information to aid data reduction. The reliability estimates tended to be low, especially for the K4 category (Inference From Principle) which tended to have only a few items or steps in each MOS, and the category scores were highly correlated with each other. We are still convinced that the Knowledge/Performance (K/P) categories provide useful information about the nature of performance as measured by these task tests, but the categories were not used in defining the basic criterion scores.

TRAINING TESTS

Criterion scores for the training knowledge tests were derived in the same way as for the job knowledge tests. The results of the expert judgments and the exploratory factor analyses suggested that the six-score solution was also a reasonable one for the training tests.

Consequently, in the subsequent analyses aimed at developing a comprehensive model of job performance, the six content categories were scored in each of the three tests (hands-on, job knowledge, school knowledge) in each MOS in Batch A.

CONCLUDING COMMENTS

Initial analyses of the data collected in the Project A Concurrent Validation were conducted for the job knowledge tests, hands-on tests, and task rating scales. The objective was to reduce large sets of task, item, and scale scores to the set of basic criterion scores that would be used with the other criterion data to develop the final criterion factor scores. Our analyses were directed at results at the task test level, at functional categories of tasks, and at behavioral requirements underlying task performance.

Analyses of the task rating scales convinced us that those scales were not reliable sources of performance data.

Factor analyses of the functional categories in each of the nine MOS led us to accept a structure of six task clusters: Communications, Vehicle Operations, Basic Soldiering Skills, Identify Targets, Safety/Survival, and Technical Tasks. Scores on these clusters of tasks, from the job knowledge and hands-on components, form the basic criterion scores.

Table 5.5

Correlations Between Criterion Factor Scores and Functional Categories for Job Knowledge Component

		Commo.	Vehicle	Basic	Identify	Survival	Technical
FACTORS							
Communications							
Vehicles		21-25					
Basic		17-51	09-48				
Identify Tgts.		09-21	12-15	10-42			
Survival		15-48	13-42	44-71	07-29		
Technical		21-56	12-65	47-63	10-32	46-76	
FUNCTIONAL CATEGORIES							
Communications		100	21-28	17-51	09-21	15-50	21-58
Vehicle Ops.		21-28	100	09-48	12-15	22-28	20-35
Navigate		12-45	06-30	65-79	12-32	26-57	31-48
Field Tech.		09-46	04-27	36-93	08-39	13-63	24-55
Weapons		12-41	10-39	67-85	04-35	37-62	34-59
Anti Air/Tank Wpns.		14	-	32	20	26	-
Customs & Laws		13-33	11-30	56-67	03-20	31-47	36-44
Identify Tgts.		09-21	12-15	10-42	100	07-32	11-33
First Aid		09-35	12-25	31-55	06-26	63-98	30-73
NBC		15-51	11-41	41-62	05-26	78-89	39-61
Technical:	13B	18-21	-	47-56	18-24	42-51	75-97
	19E	36	-	52-55	28-29	47-48	80-88
	31C	34-49	14-35	32-57	13-29	38-51	65-81
	63B	-	35-62	37-56	-	29-44	62-91
	64C	-	-	55	11	50	100
	71L	-	-	29-43	-	26-39	53-88
	91A	-	01-13	20-55	-03-19	42-76	45-98
	95B	20-31	06-20	33-53	12-17	28-46	63-85

Note: The numbers shown are the range of correlations that resulted for individual MOS; under the Technical functional category, however, the range of correlations is shown across the individual MOS Technical functional categories. Decimals have been omitted in the correlations.

Table 5.6

Correlations Between Criterion Factor Scores and Functional Categories for Hands-On Component

		Commo.	Vehicle	Basic	Survival	Technical
FACTORS						
Communications						
Vehicles		11-29				
Basic		06-26	07-15			
Survival		04-22	04-16	09-40		
Technical		06-28	07-15	12-42	10-29	
FUNCTIONAL CATEGORIES						
Communications		100	10-29	05-26	02-20	07-30
Vehicle Ops.		10-29	100	07-15	11-16	08-11
Navigate		04-21	05-13	53-100	09-35	09-24
Field Tech.		08-18	05	39-70	08-13	09-18
Weapons		-01-22	-01-14	30-85	-01-31	07-37
Anti Air/Tank Wpns.		06	-	51	12	-
Customs & Laws		-	05	46	02	-02
First Aid		06-17	04-13	05-40	67-100	04-30
NBC		-04-17	02-12	06-22	46-81	04-22
Technical:	13B	08-09	-	26-42	12-16	66-95
	19E	18-21	-	16-19	16-23	80-82
	31C	13-31	04-11	12-26	00-18	55-76
	63B	-	07-13	06-07	01-05	47-82
	64C	-	-	12	11	100
	71L	-	-	10-20	10-11	44-93
	91A	-	-	01-23	00-32	39-96
	95B	07	08	17	12	100

Note: The numbers shown are the range of correlations that resulted for individual MOS; under the Technical functional category, however, the range of correlations is shown across the individual MOS Technical functional categories. Decimals have been omitted in the correlations.

Chapter 6

DEVELOPMENT OF PERFORMANCE CRITERION SCORES FROM THE ARMY-WIDE AND MOS-SPECIFIC RATING SCALES[1]

This chapter outlines the procedures involved in developing the basic criterion scores for the Army-wide and MOS-specific performance rating data obtained in the Project A Concurrent Validation testing phase.

CONTENT OF RATINGS

Army-Wide Performance Rating Scales

The Army-wide performance scales consisted of 10 behaviorally based rating dimensions specifically developed to assess the effectiveness of first-term soldiers in the Army. Thus these scales applied to all 19 of the target MOS. Both supervisor and peer ratings of each soldier were collected using these scales.

The names of the 10 Army-wide performance dimensions are as follows:

A. Technical Knowledge/Skill
B. Effort
C. Following Regulations and Orders
D. Integrity
E. Leadership
F. Maintaining Assigned Equipment
G. Military Appearance
H. Physical Fitness
I. Self-Development
J. Self-Control

Each dimension was defined by an overall statement and contained three scaled behavioral summary statements describing different effectiveness levels. Ratings on each dimension were made on a 7-point scale ranging from 1 (Low) to 7 (High).

[1]The material in this chapter was drawn from Developing the Basic Criterion Scores for Army-Wide and MOS-Specific Ratings, by Elaine D. Pulakos and Walter C. Borman (ARI Technical Report in preparation).

Army-Wide Common Task Rating Scales

The Army-wide common task scales consisted of a set of 11 7-point rating scales, which did not employ no behavioral anchors and which included a "not observed" option. These Army-wide task dimensions were derived from the Skill Level 1 Common Task Soldier's Manual. Supervisor and peer ratings using these scales were collected only for the Batch Z MOS.

The names of the 11 common task dimensions were as follows:

A. See: Identify Threat
B. See: Estimate Range
C. Communicate: Send Radio Message
D. Navigate: In the Field
E. Shoot: Weapons Operation/Maintenance
F. Shoot: Engage Target
G. Combat: Move Under Fire
H. Combat: Camouflage Self/Equipment
I. Survive: NBC Attack
J. Survive: First Aid
K. Survive: Customs of War

MOS-Specific Performance Rating Scales

For each of the Batch A MOS, a separate set of MOS-specific, behaviorally based rating scales was developed. These instruments contained from 6 to 12 rating dimensions, each of which contained a 7-point scale ranging from 1 (Low) to 7 (High). As in the Army-wide performance rating scales, each dimension was defined by an overall statement and scaled behavioral anchors describing different levels of effectiveness. Supervisor and peer ratings of each Batch A soldier were collected using these scales.

PROCEDURES FOR DEVELOPING THE BASIC CRITERION SCORES

Two goals guided this effort. The first was to identify higher order constructs underlying the rating scale dimensions within each type of rating instrument (i.e., Army-wide performance scales, Army-wide common task scales, MOS-specific performance scales). A second objective was to identify within each instrument constructs that were similar across the different MOS and for the rating scales, constructs that were also similar across rater groups (i.e., supervisors and peers).

Samples

A total of 8,642 first-term enlisted soldiers with sufficient rating data from peers and supervisors comprised the Concurrent Validation sample. Of the total sample, 4,902 soldiers represented Batch A MOS. Recall that performance ratings of the Batch A soldiers were made using the Army-wide performance rating scales and the MOS-specific performance rating scales. The remaining 3,740 soldiers in the sample represented Batch Z MOS. Performance ratings for the Batch Z MOS were made using the Army-wide performance rating scales and the Army-wide common task rating scales. The rated sample for each of the 19 MOS is shown in Table 6.1.

Table 6.1

Concurrent Validation Sample for Army-Wide and MOS-Specific Rating Scales

MOS	Peers: Number of Ratees (Soldiers)	Peers: Total Number of Ratings	Peers: Rater/Ratee Ratio	Supervisors: Number of Ratees	Supervisors: Total Number of Ratings	Supervisors: Rater/Ratee Ratio
Batch A						
11B	679	2,377	3.50	650	1,242	1.92
13B	633	2,204	3.48	638	1,218	1.91
19E	485	1,601	3.30	490	934	1.91
31C	316	856	2.71	349	637	1.83
63B	559	1,467	2.62	597	1,158	1.94
64C	646	2,396	3.71	639	1,206	1.89
71L	422	990	2.35	460	788	1.71
91A	481	1,551	3.23	468	954	2.04
95B	681	2,543	3.73	652	1,255	1.92
Total (A)	4,902	15,985	3.26	4,943	9,392	1.90
Batch Z						
12B	684	2,325	3.40	672	1,248	1.86
16S	461	1,670	3.62	377	782	2.07
27E	141	454	3.22	143	271	1.90
51B	100	263	2.63	104	196	1.88
54E	372	1,139	3.06	372	649	1.74
55B	271	829	3.06	264	437	1.66
67N	265	867	3.27	245	421	1.72
76W	422	1,215	2.88	419	803	1.92
76Y	454	836	1.85	548	916	1.67
94B	570	1,168	2.94	546	1,030	1.89
Total (Z)	3,740	10,766	2.88	3,690	6,753	1.83
Total (A and Z)	**8,642**	**26,751**	**3.10**	**8,633**	**16,145**	**1.87**

For each soldier ratee in the sample, the goal was to obtain ratings from two supervisors and four peers who had worked with the ratee for at least two months and/or were sufficiently familiar with the ratee's job performance. The specific procedures used to identify peer and supervisor ratees can be found in Pulakos and Borman (1986a). The actual numbers of raters per ratee by MOS and type of rater (i.e., supervisor or peer) are also presented in Table 6.1. In all MOS, we came close, but did not quite achieve our goal of two supervisor and four peer ratings. However, the number of raters per ratee was sufficient to allow reasonable estimates of interrater reliability.

The appropriate rating scales were administered to groups of peer raters and separately to groups of supervisor raters. For the peer rating sessions, the groups were typically 25-35 in size. For the supervisor rating sessions, anywhere from 10 to 40 raters attended. An extremely important aspect of each rating session was a rater orientation and training program developed to reduce various types of rating errors and to persuade raters to try hard to provide accurate evaluations. All raters received this training before they made any evaluations of their peers or subordinates. The rater orientation and training program is described in detail in Pulakos and Borman (1986a).

Data Analysis

For the Army-wide performance rating scales, Army-wide common task rating scales, and MOS-specific performance rating scales, the first analyses focused on distributions of the ratings (i.e, means and standard deviations) and interrater reliabilities. These analyses were conducted within MOS and separately for the supervisor and peer raters.

Principal factor analyses were then carried out to identify the constructs underlying the dimensions in each set of rating scales. These analyses were conducted across rater groups as well as within each of the peer and supervisor rater groups. Because the job-specific measures differed for each of the nine Batch A MOS, it was necessary to factor analyze these ratings within MOS. However, factor analyses of the Army-wide measures (Army-wide performance rating scales and Army-wide common tasks scales) were performed across MOS. The results of these analyses are presented in the following section.

RESULTS

First are presented descriptive data (means, standard deviations, and interrater reliabilities) relevant to the quality of the ratings. Recall that these analyses were conducted within MOS and separately for peers and supervisors. Then, results of the factor analyses used to identify the constructs underlying the dimensions of each type of rating instrument are presented and discussed.

Rating Distributions

One criterion for assessing the quality of peer and supervisor ratings is to evaluate the means and standard deviations. Particularly in operational settings, ratings are often skewed, with most ratees receiving high performance evaluations. For-research-only administrations of rating scales (such as the present effort) often yield ratings that are more normally distributed, with lower mean ratings and greater variance in evaluations across raters.

Presented in Table 6.2 and 6.3 are the means and standard deviations of selected composite ratings for Batch A and Batch Z MOS, respectively. The data in these tables suggest that raters did not succumb to excessive leniency (overly high ratings) or restriction-in-range (rating everyone at about the same level). The dimension mean ratings, which are generally between 4 and 5 on the 7-point scales, seem reasonable in that we would expect the typical performance of a first-term soldier to be a little above average; the rationale underlying this expectation is that some percentage of poor performers will have already left the Army. Likewise, the dimension standard deviations, which are generally a little over 1.00, suggest good spread in the ratings.

Interrater Reliability

Interrater reliabilities for selected rating measures appear in Tables 6.4 and 6.5 for the Batch A and Batch Z MOS, respectively. Intraclass correlation coefficients for the Army-wide performance rating scales were encouraging, with reliabilities of the composite measures around .60. Reliabilities of the individual behavioral scales were generally lower, as would be expected.

For the MOS-specific performance rating scales, the reliabilities were somewhat lower than for the Army-wide rating scales (peer composite mdn. = .48; supervisor composite mdn. = .54), but still respectable.

Reliabilities of the Army-wide task rating scale composites were more variable across the MOS and were, in almost all cases, considerably lower than the reliabilities obtained with the behaviorally based scales. This result is probably due to the fact that the task scales contain no behavioral anchors against which the observed performance of each soldier could be compared, and the tasks being rated were specific tasks which the raters may have had varying opportunity to observe and subsequently recall.

It should be noted that the data presented in Tables 6.4 and 6.5 are intraclass correlation coefficients representing the reliabilities of mean ratings across supervisors or peers. These are the appropriate reliability estimates, because all subsequent data analyses reported here were based on mean supervisor and mean peer ratings of each ratee.

Table 6.2

Rating Dimension Means and Standard Deviations for Army-Wide and MOS-Specific Performance Rating Scales: Batch A MOS

Scales	11B	13B	19E	31C	63B	64C	71L	91A	95B
Peers									
Army-Wide Performance Rating									
Range of Dim. Means	4.02-4.94	4.08-5.02	3.97-4.77	4.28-5.07	4.03-4.85	3.99-4.87	4.40-5.20	4.05-4.86	4.23-4.89
Range of Dim. SDs	.96-1.15	.93-1.11	.87-1.11	.94-1.21	1.00-1.25	.88-1.08	.92-1.22	.88-1.22	.80-1.08
Median Dim. Means	4.54	4.57	4.48	4.76	4.54	4.58	4.91	4.61	4.74
Median Dim. SD	1.02	.99	1.03	1.04	1.10	.96	1.10	1.04	.91
MOS Performance Rating									
Range of Dim. Means	4.23-4.90	4.36-4.99	4.30-4.89	4.59-5.09	4.38-5.07	4.18-5.35	4.49-5.17	4.46-4.93	4.51-4.98
Range of Dim. SDs	.82-.98	.83-1.03	.84-1.00	.86-.97	.96-1.18	.81-.98	.93-1.06	.88-.96	.74-.93
Median Dim. Means	4.58	4.70	4.74	4.91	4.64	4.75	4.87	4.60	4.80
Median Dim. SD	.90	.93	.87	.95	1.01	.89	1.03	.94	.85
Supervisors									
Army-Wide Performance Rating									
Range of Dim. Means	3.77-5.28	3.76-5.34	3.72-5.12	3.80-5.24	3.50-5.18	3.59-4.94	4.11-5.38	3.74-4.90	4.01-5.03
Range of Dim. SDs	1.14-1.44	1.13-1.38	1.10-1.35	1.06-1.45	1.20-1.41	1.10-1.34	1.18-1.43	1.06-1.41	1.08-1.30
Median Dim. Means	4.47	4.56	4.42	4.59	4.35	4.42	4.89	4.52	4.64
Median Dim. SD	1.37	1.27	1.25	1.27	1.32	1.26	1.31	1.31	1.22
MOS Performance Rating									
Range of Dim. Means	4.00-5.09	4.18-5.13	4.13-5.07	4.24-4.82	3.97-4.89	4.09-5.31	4.42-5.34	4.11-4.57	4.42-5.01
Range of Dim. SDs	1.04-1.24	1.04-1.35	.96-1.18	.96-1.27	1.15-1.33	.97-1.24	1.09-1.33	1.00-1.20	.93-1.13
Median Dim. Means	4.56	4.69	4.72	4.71	4.18	4.75	4.92	4.49	4.82
Median Dim. SD	1.17	1.18	1.06	1.13	1.27	1.12	1.18	1.09	1.06

Table 6.3

Rating Dimension Means and Standard Deviations for Army-Wide Performance and Army-Wide Common Task Scales: Batch Z MOS

Scales	12B	16S	27E	51B	54E	55B	67N	76W	76Y	94B
<u>Peers</u>										
Army-Wide Performance Rating										
Range of Dim. Means	4.01-4.78	4.05-5.07	3.91-4.86	3.85-4.84	4.29-5.02	4.00-4.96	4.12-4.96	4.16-4.99	4.44-5.16	4.16-4.84
Range of Dim. SDs	.90-1.13	1.00-1.09	.87-1.20	1.07-1.30	.96-1.30	.94-1.25	.93-1.19	.98-1.17	1.01-1.33	.98-1.27
Median Dim. Means	4.46	4.53	4.64	4.44	4.68	4.60	4.69	4.59	4.87	4.54
Median Dim. SD	1.03	1.11	1.04	1.14	1.08	1.05	1.03	1.08	1.23	1.11
Army-Wide Common Task										
Range of Dim. Means	4.18-5.24	4.65-5.57	4.09-5.47	4.25-5.37	4.63-5.96	4.47-5.43	4.56-5.41	4.17-5.35	4.07-5.56	3.90-5.26
Range of Dim. SDs	.86-1.21	.77-1.21	.88-1.47	.82-1.25	.84-1.16	.90-1.28	.94-1.34	.95-1.36	1.06-1.48	1.07-1.50
Median Dim. Means	4.72	5.02	4.87	5.08	4.95	5.06	5.06	4.80	4.89	4.85
Median Dim. SD	.98	1.04	1.13	1.02	.97	1.04	1.12	1.09	1.34	1.25
<u>Supervisors</u>										
Army-Wide Performance Rating										
Range of Dim. Means	3.78-5.28	3.89-5.25	3.74-5.24	3.71-5.31	3.88-5.13	3.77-5.10	3.96-5.40	3.62-5.18	3.87-5.43	3.87-4.91
Range of Dim. SDs	1.09-1.42	1.18-1.38	1.08-1.38	1.18-1.51	1.19-1.46	1.23-1.53	1.21-1.39	1.15-1.34	1.12-1.44	1.15-1.45
Median Dim. Means	4.42	4.58	4.53	4.46	4.55	4.45	4.62	4.33	4.56	4.27
Median Dim. SD	1.32	1.32	1.25	1.31	1.35	1.36	1.28	1.30	1.33	1.33
Army-Wide Common Task										
Range of Dim. Means	4.17-5.58	4.62-5.62	4.11-5.56	4.13-5.65	4.38-6.02	4.12-5.50	4.29-5.61	3.85-5.29	3.95-5.74	3.53-5.04
Range of Dim. SDs	1.02-1.43	.99-1.37	.90-1.44	.93-1.47	.98-1.49	.92-1.49	1.00-1.50	1.04-1.42	1.15-1.56	1.08-1.58
Median Dim. Means	4.79	4.93	4.95	4.80	4.81	4.71	4.80	4.60	4.66	4.24
Median Dim. SD	1.29	1.20	1.00	1.17	1.17	1.24	1.29	1.20	1.34	1.32

Table 6.4

Interrater Reliabilities for Army-Wide and MOS-Specific Composite Performance Rating Scales: Batch A MOS

Scale	11B	13B	19E	31C	63B	64C	71L	91A	95B	Batch A
Peers										
Army-Wide Composite	.63	.61	.62	.57	.54	.61	.41	.59	.63	.58
Scale Range	.41- .61	.35- .60	.38- .63	.30- .61	.27- .54	.40- .60	.18- .62	.36- .66	.35- .71	
Median	.54	.49	.47	.44	.48	.49	.34	.41	.52	
MOS-Specific Composite	.57	.49	.54	.47	.47	.48	.30	.40	.47	.42
Scale Range	.19- .50	.29- .80	.26- .45	.13- .42	.15- .45	.27- .45	.00- .26	.13- .39	.33- .54	
Median	.42	.34	.41	.34	.33	.35	.11	.26	.42	
Supervisors										
Army-Wide Composite	.69	.54	.64	.64	.65	.65	.72	.68	.60	.65
Scale Range	.48- .61	.36- .59	.45- .59	.40- .64	.48- .69	.41- .62	.40- .68	.51- .62	.33- .61	
Median	.58	.48	.51	.49	.57	.51	.54	.56	.49	
MOS-Specific Composite	.59	.49	.57	.58	.65	.44	.58	.51	.48	.54
Scale Range	.35- .52	.28- .45	.30- .48	.31- .51	.40- .62	.27- .45	.31- .54	.17- .57	.25- .42	
Median	.42	.40	.40	.42	.48	.34	.45	.34	.32	

NOTE: Negative intraclass correlations were reported as .00 because negative ICCs are uninterpretable. Thus, the reliabilities of .00 indicate that the within variance was equal to or greater than the across variance.

Table 6.5

Interrater Reliabilities for Army-Wide Performance and Army-Wide Common Task Composite Rating Scales: Batch Z MOS

Scale	12B	16S	27E	51B	54E	55B	67N	76W	76Y	94B	Batch Z
<u>Peers</u>											
Army-Wide Composite	.65	.69	.71	.67	.64	.45	.68	.46	.33	.53	.58
Scale Range	.43-.61	.48-.66	.47-.68	.35-.69	.39-.67	.15-.48	.49-.68	.28-.50	.05-.44	.24-.50	
Median	.54	.58	.51	.55	.53	.39	.55	.39	.22	.43	
Army-Wide Common Composite	.49	.52	.66	.44	.46	.30	.40	.27	.16	.29	.40
Scale Range	.19-.45	.20-.58	.39-.62	.19-.50	.22-.46	.00-.36	.16-.45	.11-.33	.00-.33	.06-.28	
Median	.37	.31	.46	.33	.34	.19	.32	.23	.16	.15	
<u>Supervisors</u>											
Army-Wide Composite	.61	.68	.70	.81	.66	.66	.59	.56	.56	.60	.64
Scale Range	.40-.61	.46-.66	.36-.65	.46-.75	.42-.65	.36-.61	.36-.61	.35-.55	.38-.57	.42-.53	
Median	.49	.57	.55	.61	.53	.52	.48	.45	.45	.50	
Army-Wide Common Composite	.24	.52	.61	.67	.45	.35	.44	.49	.40	.18	.44
Scale Range	.00-.33	.24-.52	.00-.61	.28-.82	.00-.52	.00-.42	.00-.73	.16-.47	.18-.56	.05-.23	
Median	.18	.33	.39	.44	.40	.00	.35	.43	.37	.14	

NOTE: Negative intraclass correlations were reported as .00 because negative ICCs are uninterpretable. Thus, the reliabilities of .00 indicate that the within variance was equal to or greater than the across variance.

Factor Analyses Results for the Rating Scales

Army-Wide Performance Rating Scales. Principal factor analyses with a varimax rotation were employed to identify the constructs underlying the rating scale dimensions. For the Army-wide performance rating scales, these analyses were performed across MOS for peer raters, for supervisor raters, and for the combined peer and supervisor rater groups. Virtually identical results were obtained for all three rater groups, and a three-factor solution was chosen as most psychologically meaningful. The names of the factors and the rating dimensions loading highest on each factor are shown in Table 6.6. Loadings for the rotated factor solutions for peers, supervisors, and the combined group are shown in Tables 6.7, 6.8, and 6.9, respectively.

To determine how well the factor solutions based on all 19 MOS would hold up within individual MOS, we first computed factor scores using the factor scoring matrixes generated from the analyses across MOS. Thus, factor scores were computed within the peer rater group, within the supervisor rater group, and for the combined peer and supervisor rater group. Then, within rater groups (peer, supervisor, combined) and for each of the 19 MOS as well as across the 19 MOS, we computed correlations between the factor scores and the original behavioral dimension ratings.

Correlations between the factor scores and the Army-wide dimension ratings are shown in Tables 6.10, 6.11, and 6.12 for peers, supervisors, and the combined group, respectively. The results of these analyses generally supported the stability and appropriateness of the three-factor structure across rating source and MOS. Correlations between dimension ratings and factor scores for each rating source-by-MOS combination were checked to identify instances in which dimension ratings related higher with a factor other than the one they were supposed to correlate highest with according to the across-MOS results.

For peer ratings, Maintaining Equipment shifted back and forth between Factors 1 and 3. For 7 of the 19 MOS, correlations between that dimension's ratings and Factor 3 were higher than the correlations with Factor 1. Conceptually, this is not too troublesome because Maintaining Equipment, rather than being seen as a core technical skill or motivation-related dimension (Factor 1), might be seen as a more peripheral, appearance- or maintenance-oriented dimension (Factor 3). In addition, for 2 of the 19 MOS, ratings on Self-Development correlated higher with Factor 3 than with Factor 1.

For the supervisor raters, four MOS had ratings on Maintaining Equipment that correlated higher on Factor 3 than on Factor 1, and one MOS had Self-Development correlating as high with Factor 3 as with Factor 1. When ratings from the two rating sources were pooled (Table 6.12), ratings for only two MOS on Maintaining Equipment correlated higher with Factor 3 than with Factor 1.

Table 6.6

Army-Wide Performance Rating Scales Factors

Factor 1:	Job-Relevant Skills and Motivation Technical Knowledge/Skill Leadership Effort Self-Development Maintaining Equipment
Factor 2:	Personal Discipline Following Regulations Self-Control Integrity
Factor 3:	Physical Fitness and Military Bearing Military Appearance Physical Fitness

Table 6.7

Army-Wide Performance Rating Scales Three-Factor Solution for Peer Raters

Rotated Factor Pattern[a]			
Factor 1	Factor 2	Factor 3	Dimensions
.65	.30	.32	A: Technical Skill
.62	.32	.40	E: Leadership
.60	.45	.29	B: Effort
.49	.40	.38	I: Self-Development
.43	.35	.39	F: Maintaining Equipment
.34	.65	.28	C: Following Regulations
.20	.57	.19	J: Self-Control
.43	.55	.29	D: Integrity
.28	.31	.56	G: Military Appearance
.22	.15	.50	H: Physical Fitness

[a]Factor 1 - Job-Relevant Skills and Motivation; Factor 2 - Personal Discipline; Factor 3 - Physical Fitness and Military Bearing.

Table 6.8

Army-Wide Performance Rating Scales Three-Factor Solution for Supervisor Raters

Rotated Factor Pattern[a]

Factor 1	Factor 2	Factor 3	Dimensions
.70	.40	.25	B: Effort
.69	.26	.29	A: Technical Skill
.68	.30	.34	E: Leadership
.55	.34	.39	I: Self-Development
.53	.32	.38	F: Maintaining Equipment
.42	.66	.30	C: Following Regulations
.22	.61	.23	J: Self-Control
.49	.57	.29	D: Integrity
.32	.32	.55	G: Military Appearance
.19	.17	.47	H: Physical Fitness

[a]Factor 1 - Job-Relevant Skills and Motivation; Factor 2 - Personal Discipline; Factor 3 - Physical Fitness and Military Bearing.

Table 6.9

Army-Wide Performance Rating Scales Three-Factor Solution for Combined Peer and Supervisor Raters

Rotated Factor Pattern[a]

Factor 1	Factor 2	Factor 3	Dimensions
.71	.28	.30	A: Technical Skill
.69	.30	.37	E: Leadership
.69	.43	.26	B: Effort
.57	.38	.38	I: Self-Development
.54	.34	.35	F: Maintaining Equipment
.41	.69	.30	C: Following Regulations
.22	.63	.20	J: Self-Control
.50	.59	.28	D: Integrity
.32	.32	.57	G: Military Appearance
.21	.15	.49	H: Physical Fitness

[a]Factor 1 - Job-Relevant Skills and Motivation; Factor 2 - Personal Discipline; Factor 3 - Physical Fitness and Military Bearing.

Table 6.10

Correlations Between Army-Wide Performance Factor[a] Scores and Army-Wide Performance Rating Dimensions for Peer Raters

	Factor 1	Factor 2	Factor 3
Factor 1			
Factor 2	.39 (.23-.51)		
Factor 3	.48 (.36-.59)	.34 (.16-.51)	
Technical Knowledge	.87 (.83-.91)	.40 (.27-.50)	.50 (.36-.61)
Leadership	.83 (.79-.88)	.43 (.31-.52)	.61 (.53-.70)
Effort	.81 (.74-.86)	.61 (.50-.71)	.44 (.35-.56)
Self-Development	.66 (.54-.77)	.55 (.35-.65)	.60 (.32-.75)
Maintaining Equipment	.59 (.42-.69)	.48 (.37-.60)	.61 (.53-.68)
Following Regulations	.46 (.35-.60)	.89 (.85-.91)	.43 (.28-.56)
Self-Control	.27 (.04-.36)	.78 (.72-.84)	.29 (.20-.39)
Integrity	.58 (.47-.67)	.75 (.72-.81)	.45 (.27-.59)
Military Appearance	.38 (.18-.51)	.42 (.30-.56)	.87 (.83-.90)
Physical Fitness	.30 (.12-.40)	.21 (-.02-.42)	.77 (.71-.81)

Note: In parentheses is the range of correlations that resulted for individual MOS. Above the numbers in parentheses are correlations for the entire sample, combining MOS.

[a]Factor 1 - Job-Relevant Skills and Motivation; Factor 2 - Personal Discipline; Factor 3 - Physical Fitness and Military Bearing.

Table 6.11

Correlations Between Army-Wide Performance Factor[a] Scores and Army-Wide Performance Rating Dimensions for Supervisor Raters

	Factor 1	Factor 2	Factor 3
Factor 1			
Factor 2	.30 (.13-.36)		
Factor 3	.39 (.27-.51)	.34 (.23-.47)	
Technical Knowledge	.86 (.83-.89)	.35 (.19-.42)	.46 (.34-.54)
Leadership	.84 (.81-.92)	.40 (.22-.49)	.53 (.44-.62)
Effort	.87 (.83-.89)	.53 (.40-.59)	.39 (.25-.48)
Self-Development	.67 (.61-.73)	.45 (.27-.50)	.62 (.55-.68)
Maintaining Equipment	.66 (.54-.71)	.42 (.30-.52)	.60 (.51-.72)
Following Regulations	.52 (.41-.58)	.88 (.86-.90)	.48 (.38-.56)
Self-Control	.27 (.13-.36)	.81 (.78-.84)	.36 (.27-.44)
Integrity	.60 (.48-.65)	.76 (.63-.80)	.47 (.34-.57)
Military Appearance	.40 (.31-.50)	.43 (.28-.52)	.87 (.84-.90)
Physical Fitness	.24 (.15-.36)	.22 (.12-.37)	.74 (.68-.81)

Note: In parentheses is the range of correlations that resulted for individual MOS. Above the numbers in parentheses are correlations for the entire sample, combining MOS.

[a]Factor 1 - Job-Relevant Skills and Motivation; Factor 2 - Personal Discipline; Factor 3 - Physical Fitness and Military Bearing.

Table 6.12

Correlations Between Army-Wide Performance Factor[a] Scores and Army-Wide Performance Rating Dimensions for Combined Peer and Supervisor Raters

	Factor 1	Factor 2	Factor 3
Factor 1			
Factor 2	.28 (.13-.37)		
Factor 3	.40 (.30-.51)	.28 (.13-.36)	
Technical Knowledge	.88 (.85-.91)	.36 (.25-.45)	.46 (.32-.58)
Leadership	.85 (.82-.89)	.38 (.27-.47)	.58 (.50-.64)
Effort	.85 (.80-.88)	.55 (.44-.65)	.41 (.31-.49)
Self-Development	.70 (.63-.79)	.48 (.36-.55)	.59 (.47-.75)
Maintaining Equipment	.67 (.59-.74)	.43 (.36-.54)	.55 (.47-.67)
Following Regulations	.50 (.41-.63)	.89 (.86-.91)	.46 (.36-.56)
Self-Control	.27 (.08-.35)	.81 (.79-.85)	.31 (.15-.36)
Integrity	.61 (.46-.70)	.76 (.71-.81)	.44 (.34-.53)
Military Appearance	.39 (.23-.49)	.41 (.32-.53)	.89 (.75-.92)
Physical Fitness	.25 (.14-.37)	.18 (-.03-.31)	.75 (.69-.80)

Note: In parentheses is the range of correlations that resulted for individual MOS. Above the numbers in parentheses are correlations for the entire sample, combining MOS.

[a]Factor 1 - Job-Relevant Skills and Motivation; Factor 2 - Personal Discipline; Factor 3 - Physical Fitness and Military Bearing.

No other such "reversals" occurred for individual rating sources and MOS. For the vast majority of dimension-by-MOS combinations, the dimensions correlated highest with their target factors as defined by the factor analysis results for the entire sample. This admittedly is only one possible way to explore stability of the three-factor solution across MOS and rating source, but the results do suggest that the three-factor solution represents a consistent, interpretable summary of the 10 Army-wide dimensions.

Army-Wide Common Task Rating Scales. Principal factor analyses with a varimax rotation were again utilized to identify the constructs underlying the task dimensions. These analyses were also conducted across the 10 Batch Z MOS for peers, supervisors, and the combined rater group. Recall that only Batch Z raters completed these scales. A three-factor solution was chosen as the most psychologically meaningful, although there was some small variation between the three rating groups in terms of which dimensions loaded highest on each factor.

The names of the factors and the task dimensions loading highest on each are shown in Table 6.13 for the three rater groups. As can be seen in the table, the solutions are identical across the rater groups with the exception of two task dimensions (Combat: Move Under Fire and Combat: Camouflage Self/Equipment). The rotated factor loading matrixes for peers, supervisors, and the combined rater group are shown in Tables 6.14, 6.15, and 6.16, respectively.

As in the analysis plan for the Army-wide performance rating scales, an objective in analyzing the Army-wide common task scales was to determine whether the factor solutions based on all 10 MOS would hold up within rating source and individual MOS. Thus, for each of the rater groups (peers, supervisors, and the two groups combined), we computed correlations within MOS between Table 6.13 the original task dimension ratings and task factor scores. Within each rater group, the three-factor solutions held up reasonably well, although in five MOS the two Combat dimensions correlated higher with a factor other than their "target" factor. The correlations between the original task dimensions and the factor scores are presented in Tables 6.17, 6.18, and 6.19 for peers, supervisors, and combined peers and supervisors, respectively.

Because the two Combat dimensions (i.e., Camouflage Self/Equipment and Move Under Fire) did not always correlate highest with their target factor within individual MOS and also loaded on different factors for the three rater groups, we decided to select the most psychologically meaningful solution and to use that solution as a basis for calculating the common task rating scores for all MOS and rater groups. The solution that resulted for the supervisor raters seemed to make the most sense conceptually and was thus chosen as our final solution. Accordingly, Combat: Move Under Fire was included in Factor 1 (Field Skills), while Combat: Camouflage Self/Equipment was included in Factor 2 (Weapons Operation and Maintenance).

Table 6.13

Army-Wide Common Task Rating Scales Factors

Factor 1: Field Skills	Factor 2: Weapons Operation & Maintenance	Factor 3: Survival
Peers		
Navigate: In the Field	Combat: Move Under Fire	Combat: Camouflage Self/Equip
Communicate: Send Radio Message	Shoot: Weapons Oper/Maint	Survive: First Aid
See: Identify Threat	Shoot: Engage Target	Survive: Customs of War
See: Estimate Range		Survive: NBC Attack
Supervisors		
Navigate: In the Field	Shoot: Weapons Oper/Maint	Survive: First Aid
Communicate: Send Radio Message	Shoot: Engage Target	Survive: Customs of War
See: Identify Threat	Shoot: Camouflage Self/Equip	Survive: NBC Attach
See: Estimate Range		
Combat: Move Under Fire		
Combined Peers and Supervisors		
Navigate: In the Field	Shoot: Weapons Oper/Maint	Combat: Camouflage Self/Equip
Communicate: Send Radio Message	Shoot: Engage Target	Survive: First Aid
See: Identify Threat		Survive: Customs of War
See: Estimate Range		Survive: NBC Attack
Combat: Move Under Fire		

Table 6.14

Army-Wide Common Task Ratings Three-Factor Solution for Peer Raters

Rotated Factor Pattern[a]				
Factor 1	Factor 2	Factor 3	Dimensions[b]	
.59	.33	.28	Comm:	Send Radio Message
.58	.33	.30	Nav:	In the Field
.55	.35	.28	See:	Estimate Range
.55	.22	.24	See:	ID Threat
.46	.24	.44	Surv:	Customs of War
.28	.63	.23	Shoot:	Engage Target
.32	.60	.28	Shoot:	Weapon Oper Maint
.37	.47	.33	Comb:	Move Under Fire
.40	.34	.52	Surv:	First Aid
.33	.33	.50	Surv:	NBC Attack
.30	.45	.47	Comb:	Camouflage Self/Equip

[a]Factor 1 - Field Skills; Factor 2 - Weapons Operation and Maintenance; Factor 3 - Survival.
[b]Comm - Communicate; Nav - Navigate; Surv - Survive; Comb - Combat.

Table 6.15

Army-Wide Common Task Ratings Three-Factor Solution for Supervisor Raters

Rotated Factor Pattern[a]			
Factor 1	Factor 2	Factor 3	Dimensions[b]
.65	.34	.22	Nav: In the Field
.63	.25	.23	Comm: Send Radio Message
.60	.21	.28	See: ID Threat
.60	.34	.25	See: Estimate Range
.52	.37	.31	Comb: Move Under Fire
.29	.69	.24	Shoot: Weapon Oper Maint
.28	.66	.18	Shoot: Engage Target
.29	.54	.38	Comb: Camouflage Self/Equip
.36	.43	.57	Surv: First Aid
.44	.23	.48	Surv: Customs of War
.35	.37	.48	Surv: NBC Attack

[a]Factor 1 - Field Skills; Factor 2 - Weapons Operation and Maintenance; Factor 3 - Survival.
[b]Nav - Navigate; Comm - Communicate; Comb - Combat; Surv - Survive.

Table 6.16

Army-Wide Common Task Ratings Three-Factor Solution for Combined Peer and Supervisor Raters

Rotated Factor Pattern[a]			
Factor 1	Factor 2	Factor 3	Dimensions[b]
.60	.32	.32	Nav: In the Field
.60	.27	.29	Comm: Send Radio Message
.59	.33	.27	See: Estimate Range
.56	.23	.26	See: ID Threat
.43	.39	.30	Comb: Move Under Fire
.28	.64	.22	Shoot: Engage Target
.30	.63	.29	Shoot: Weapon Oper Maint
.32	.45	.43	Comb: Camouflage Self/Equip
.38	.34	.55	Surv: First Aid
.33	.34	.57	Surv: NBC Attack
.42	.23	.47	Surv: Customs of War

[a]Factor 1 - Field Skills; Factor 2 - Weapons Operation and Maintenance; Factor 3 - Survival.
[b]Nav - Navigate; Comm - Communicate; Comb - Combat; Surv - Survive.

Table 6.17

Correlations Between Army-Wide Common Task Factor[a] Scores and Army-Wide Common Task Dimensions for Peer Raters

	Factor 1	Factor 2	Factor 3
Factor 1			
Factor 2	.37 (-.28-.62)		
Factor 3	.48 (.38-.61)	.45 (.21-.58)	
Navigate: In the Field	**.79 (.66-.86)**	.46 (.39-.70)	.48 (.31-.60)
Communicate: Send Radio Message	**.80 (.62-.86)**	.46 (.39-.66)	.44 (.30-.62)
See: Identify Threat	**.75 (.61-.83)**	.31 (.15-.54)	.38 (.27-.54)
See: Estimate Range	**.75 (.70-.83)**	.48 (.37-.64)	.44 (.37-.64)
Combat: Move Under Fire	.50 (.43-.67)	**.66 (.51-.77)**	.52 (.46-.68)
Shoot: Weapons Operation/Maint	.44 (.32-.61)	**.84 (.72-.91)**	.45 (.25-.64)
Shoot: Engage Target	.38 (.26-.66)	**.88 (.75-.93)**	.37 (.11-.57)
Combat: Camouflage Self/Equipment	.40 (.35-.57)	.63 (.48-.71)	**.75 (.63-.84)**
Survive: First Aid	.55 (.39-.73)	.47 (.37-.63)	**.82 (.77-.88)**
Survive: Customs of War	.62 (.57-.78)	.33 (.24-.47)	**.70 (.61-.83)**
Survive: NBC Attack	.44 (.32-.69)	.46 (.16-.62)	**.78 (.72-.86)**

Note: In parentheses is the range of correlations that resulted for individual MOS. Above the numbers in parentheses are correlations for the entire sample, combining MOS.

[a]Factor 1 - Field Skills; Factor 2 - Weapons Operation and Maintenance; Factor 3 - Survival.

Table 6.18

Correlations Between Army-Wide Common Task Factor[a] Scores and Army-Wide Common Task Dimensions for Supervisor Raters

	Factor 1	Factor 2	Factor 3
Factor 1			
Factor 2	.25 (-.20-.61)		
Factor 3	.39 (.13-.71)	.37 (-.02-.59)	
Navigate: In the Field	$\boxed{.82\ (.75-.88)}$	.44 (.08-.75)	.39 (.10-.63)
Communicate: Send Radio Message	$\boxed{.80\ (.72-.88)}$	.32 (-.03-.65)	.37 (.08-.62)
See: Identify Threat	$\boxed{.75\ (.71-.83)}$	.27 (-.30-.56)	.44 (.25-.62)
See: Estimate Range	$\boxed{.75\ (.67-.88)}$	.43 (.06-.77)	.40 (.32-.57)
Combat: Move Under Fire	$\boxed{.66\ (.57-.78)}$	.47 (.24-.69)	.49 (.25-.68)
Shoot: Weapons Operation/Maint	.37 (-.04-.73)	$\boxed{.89\ (.81-.92)}$	.38 (-.07-.55)
Shoot: Engage Target	.35 (-.07-.62)	$\boxed{.86\ (.81-.90)}$	.28 (-.10-.43)
Combat: Camouflage Self/Equipment	.37 (.06-.60)	$\boxed{.70\ (.35-.81)}$	.60 (.51-.78)
Survive: First Aid	.45 (.22-.65)	.56 (.11-.76)	$\boxed{.80\ (.73-.86)}$
Survive: Customs of War	.55 (.35-.69)	.30 (.04-.52)	$\boxed{.77\ (.61-.87)}$
Survive: NBC Attack	.44 (.26-.72)	.48 (.32-.72)	$\boxed{.76\ (.56-.84)}$

Note: In parentheses is the range of correlations that resulted for individual MOS. Above the numbers in parentheses are correlations for the entire sample, combining MOS.

[a]Factor 1 - Field Skills; Factor 2 - Weapons Operation and Maintenance; Factor 3 - Survival.

Table 6.19

Correlations Between Army-Wide Common Task Factor[a] Scores and Army-Wide Common Task Dimensions for Combined Peer and Supervisor Raters

	Factor 1	Factor 2	Factor 3
Factor 1			
Factor 2	.32 (.17-.55)		
Factor 3	.47 (.40-.62)	.41 (.34-.58)	
Navigate: In the Field	$\boxed{.80\ (.69\text{-}.87)}$	.44 (.30-.65)	.49 (.40-.62)
Communicate: Send Radio Message	$\boxed{.80\ (.65\text{-}.86)}$	.37 (.22-.54)	.44 (.38-.57)
See: Identify Threat	$\boxed{.74\ (.65\text{-}.79)}$	.37 (.09-.52)	.44 (.31-.54)
See: Estimate Range	$\boxed{.78\ (.75\text{-}.85)}$	.46 (.37-.64)	.42 (.32-.63)
Combat: Move Under Fire	$\boxed{.57\ (.50\text{-}.70)}$	.54 (.41-.65)	.47 (.37-.66)
Shoot: Weapons Operation/Maint	.40 (.24-.59)	$\boxed{.87\ (.80\text{-}.89)}$	.45 (.34-.65)
Shoot: Engage Target	.37 (.17-.55)	$\boxed{.88\ (.82\text{-}.91)}$	.34 (.18-.51)
Combat: Camouflage Self/Equipment	.41 (.26-.56)	.63 (.48-.68)	$\boxed{.67\ (.59\text{-}.74)}$
Survive: First Aid	.51 (.39-.63)	.46 (.33-.66)	$\boxed{.85\ (.81\text{-}.89)}$
Survive: Customs of War	.56 (.50-.73)	.32 (.24-.48)	$\boxed{.73\ (.63\text{-}.79)}$
Survive: NBC Attack	.44 (.34-.63)	.47 (.38-.55)	$\boxed{.79\ (.70\text{-}.87)}$

Note: In parentheses is the range of correlations that resulted for individual MOS. Above the numbers in parentheses are correlations for the entire sample, combining MOS.

[a]Factor 1 - Field Skills; Factor 2 - Weapons Operation and Maintenance; Factor 3 - Survival.

MOS-Specific Performance Rating Scales. For the MOS-specific performance rating scales, principal factor analyses with a varimax rotation were conducted within MOS and separately for the peer and supervisor raters. Recall that one objective in developing the basic criterion scores was to look for common themes that might be evident across MOS, even though different dimensions comprised each of the nine sets of MOS-specific performance rating scales.

Inspection of the factor analyses revealed a similar two-factor solution that could be used for all nine MOS. In particular, the rating dimensions loading highest on one of the factors consisted mainly of core job requirements and tasks, while those loading highest on the second factor were more peripheral (as opposed to core) job duties. Accordingly, for all MOS, a two-factor solution was chosen to represent the MOS-specific aspect of the criterion domain, with the factors named as follows: 1 - Core Responsibilities, and 2 - Other Responsibilities.

It should be mentioned that within some MOS, slightly different two-factor solutions resulted for peer versus supervisor raters. Under such circumstances, the most psychologically meaningful solution was selected. The rating dimensions loading highest on each factor within each MOS are shown in Table 6.20.

COMBAT EFFECTIVENESS RATINGS

The combat effectiveness rating scales were summated scales based on the 40 items that survived the field tests; that is, the same items were used for each MOS. This scale, like the Army-wide rating scales, was intended to be appropriate for any MOS. It is the only criterion that specifically addresses combat performance for all Project A MOS, and is expressly designed to evaluate performance under degraded conditions and the increased confusion, workload, and uncertainty of a combat environment (Campbell, 1987).

A factor analysis of these items based on the combined samples from the Concurrent Validation suggested that only two factors could be extracted. The first factor contained items that seemed to reflect performance under adverse, difficult, or dangerous conditions. The second was composed largely of items dealing with making mistakes, getting into trouble, or creating discipline problems. Consequently, items within each factor were summed to produce two scores for expected combat effectiveness, Performing Under Adverse Conditions and Avoiding Mistakes.

Table 6.20

MOS-Specific Performance Rating Scales Factors

MOS 11B - Infantryman	
1. Core Responsibilities	2. Other Responsibilities
Reconnaissance and Patrol	Navigation
Avoid Enemy Detection	Maintaining Supply/Equip/Weap
Courage/Proficiency in Battle	Assist/Lead Others
Fighting Position	Guard/Security Duties
Prisoners of War	Field Sanitation
Operate Radio	
Use of Weapons/Equipment	
MOS 13B - Cannon Crewman	
1. Core Responsibilities	2. Other Responsibilities
Load/Unload Howitzer	Prepare Howitzer
Gunnery	Drive/Maintain Vehicles
Receive/Relay Communications	Loading Out Equipment
Record Keeping	Prepare Ammo for Fire
Position Improvement	Setting Up Communications
MOS 19E - Armor Crewman	
1. Core Responsibilities	2. Other Responsibilities
Maintain Tank	Store Ammo Aboard Tanks
Drive/Recover Tanks	Load/Unload Guns
Engage Targets with Tank Gun	Maintain Guns
Operate/Maintain Comm Equip	
Prepare Tank for Field	
MOS 31C - Single Channel Radio Operator	
1. Core Responsibilities	2. Other Responsibilities
Inspect/Service Equipment	Prepare Reports
Install/Repair Equipment	Maintain Security
Operate Communication Device	Provide Safe Transportation
MOS 63[a] - Light Wheel Vehicle Mechanic	
1. Core Responsibilities	2. Other Responsibilities
Inspect/Test with Equip	Safety Mindedness
Troubleshooting	Determine Task Requirements
Repair	Administrative Duties
Perform Routine Maint	Vehicle Operation
Recovery	Use Technical Documents

(Continued)

Table 6.20 (Continued)

MOS-Specific Performance Rating Scales Factors

MOS 64C - Motor Transport Operator	
1. Core Responsibilities	2. Other Responsibilities
Safety Mindedness	Self-Recover Vehicles
Park/Secure Vehicles	Use Maps/Follow Routes
Drive Vehicles	Vehicle Coupling
Check/Maintain Vehicles	Perform Dispatcher Duties
Load Cargo/Transp Personnel	
Perform Admin Duties	

MOS 71L - Administrative Specialist	
1. Core Responsibilities	2. Other Responsibilities
Distribute Documents	Keep Records
Prepare Documents	Establish Files IAW TAFFS
Maintain Office Resources	Post Regulations
Provide Customer Service	Safeguard Classified Documents

MOS 91A - Medical Specialist	
1. Core Responsibilities	2. Other Responsibilities
Keep Medical Records	Prepare/Inspect Field Clinic
Provide Patient Care	Arrange Transp Injured
Provide Health Care to Army	Maint/Oper Med Vehicles
Maint Med Supply and Equip	Respond to Emergencies
Dispense Medication	

MOS 95B - Military Police	
1. Core Responsibilities	2. Other Responsibilities
Investigate Crime	Respond to Emergencies
Patrolling	Courage in Battle
Traffic Control	Avoid Enemy Detection
Communication Skills	Navigation
Promote Public Image	Use of Weapons
Provide Security	

[a] The MOS 63B dimension titled Use Tools/Test Equipment was not included in either factor.

CONCLUDING COMMENTS

The principal objective of this chapter was to describe the development of the basic criterion scores for various types of rating measures that were collected as part of the Project A large-scale Concurrent Validation effort. Identification of the criterion scores was accomplished through the use of principal factors analyses within each rating instrument.

To summarize the results of these analyses:

- A three-factor solution (Job-Relevant Skills and Motivation, Personal Discipline, and Physical Fitness and Military Bearing) was chosen as the most psychologically meaningful for the Army-wide performance rating scales.

- For the Army-wide common task rating scales, a three-factor solution also resulted (Field Skills, Weapon Operation and Maintenance, and Survival).

- Factor analyses of the MOS-specific rating scales yielded a two-factor solution across all nine MOS (Core Responsibilities, and Other Responsibilities).

- Factor analysis of the combat rating scales, using the combined sample, also produced a two-factor solution (Performing Under Adverse Conditions and Avoiding Mistakes).

Chapter 7

THE MODELING OF CRITERION PERFORMANCE AND DEVELOPMENT OF CRITERION FACTOR SCORES[1]

This chapter recounts a series of activities directed toward modeling job performance in the Project A population of jobs and maximizing our understanding of the criterion measures described in the earlier chapters.

THE INITIAL FRAMEWORK

Overall, the criterion development work in Project A was guided by a particular "theory" of performance, which may be briefly stated as follows: Job performance really is multidimensional; there is not one outcome, one factor, or one anything that can be pointed to and labeled as job performance; it is manifested by a wide variety of behaviors, or things people do, that are judged to be important for accomplishing the goals of the organization.

Two General Factors

For the population of entry-level enlisted positions, we postulated that there are two major types of job performance components. The first is composed of components that are *specific to* a particular job; that is, measures of such components would reflect specific technical competence or specific job behaviors that are not required for other jobs. The second type is composed of components that are defined and measured in the same way for every job; these are referred to as Army-wide criterion factors.

For the job-specific components, we anticipated that there would be a relatively small number of distinguishable factors of technical performance that would be a function of different abilities or skills and that would be reflected by different task content.

The Army-wide concept incorporates the basic idea that total performance is much more than task or technical proficiency. It might include such attributes as contributions to teamwork, continuing self-development, support for the norms and customs of the organization, and perseverance in the face of adversity.

---

[1]This chapter is based on material from two papers: (a) Project A: When the Textbook Goes Operational, by John P. Campbell; (b) A Latent Structure Model of Job Performance Factors, by Lauress L. Wise, John P. Campbell, Jeffrey J. McHenry, and Lawrence M. Hanser. Both papers were presented at the 1986 Annual Convention of the American Psychological Association in Washington and are available in the ARI Research Note which supplements this report.

In sum, the working model of total performance with which the project began viewed performance as multidimensional within the two broad categories of factors. The job analysis and criterion construction methods were designed to "discover" the content of these factors by means of an exhaustive description of the total performance domain, several iterations of data collection, and the use of multiple methods for identifying basic performance factors.

Factors vs. a Composite

Saying that performance is multidimensional does not preclude using just one index of an individual's contributions in order to make a specific personnel decision (e.g., select/not select, promote/not promote). As argued by Schmidt and Kaplan (1971) some years ago, it seems quite reasonable for the organization to scale the importance of each major performance factor relative to a particular personnel decision that must be made, and to combine the weighted factor scores into a composite that represents the total contribution or utility of an individual's performance, within the context of that decision. That is, the way in which performance information is weighted and combined is a value judgment by the organization. Determining the specific combinational rules (e.g., simple sum, weighted sum, nonlinear combination) that best reflect what the organization is trying to accomplish is a matter for research.

Needed: The Latent Structure of Performance

If all the Project A rating scales were used separately for each type of rater and the MOS-specific measures were aggregated at the task or instructional module level, there would be approximately 200 criterion scores on each individual--too many to handle. Adding them all up into a composite is a bit too atheoretical, and developing a reliable and homogeneous measure of the general factor violates the basic notion that performance is multidimensional. A more formal way to model performance is to think in terms of its latent structure, postulate what that might be, and then resort to a confirmatory analysis.

Unfortunately, it is true that we simply know a lot more about predictor constructs than we do about job performance constructs. There are volumes of research on the former, and almost none on the latter. For personnel psychologists it is almost second nature to talk about predictors in terms of theories and constructs. However, on the performance side, the textbooks are virtually silent. Only a few people have even raised the issue (e.g., Dunnette, 1963; Wallace, 1965).

Given this initial disparity, we used the expert judgment of the Project A staff, the previous literature, and data from pilot and field tests to formulate a target model. In the field tests, the various versions of the criterion measures were administered to 100-150 people from each of nine MOS. These data and the development work leading up to them are summarized in Campbell (1987) and Campbell and Harris (1985). A picture we drew at the time is shown in Figure 7.1; it is included only to show one stage in the almost continuous process of bootstrapping ourselves toward a more final conceptual description of the predictor/criterion space.

The target model was then subjected to what might be described as "quasi" confirmatory analysis, using data from the Concurrent Validation sample. The purpose was to consider whether a single model of the latent structure of job performance would fit the data for all nine jobs. It is the result from these analyses that we report here.

PROCEDURE

As described previously, the criterion measures were administered to a Concurrent Validation sample of 400-600 people in each of the 19 jobs (MOS). The complete array of performance measures is repeated in Figure 7.2.

Previous chapters have described how each of the major sets of criterion measures was reduced from a large number of item, task, or individual scale scores to a smaller set of factor or category scores. The results of this first level of aggregation have been referred to as the "basic" array of criterion scores. A summary of these results follows.

Reduction of the Hands-On and Written Test Variables

Tasks were grouped into "functional or content categories" on the basis of the similarity in task content. The 30 tasks sampled for each job were clustered into 8 to 15 categories. Each of the school knowledge items was similarly grouped into a specific content category.

Ten of the categories were common to some or all of the jobs (e.g., first aid, basic weapons, field techniques). Each job, except Infantryman, also had two to five performance categories that were unique, or job specific.

Next, scores were computed for each content category within each of the three sets of measures. For the hands-on test, the functional category score was the mean percentage of successfully completed steps across all of the tasks assigned to that category. For the job knowledge test and the school knowledge test, the functional category score was the percentage of items within that category that were answered correctly.

After category scores were computed, they were factor analyzed via principal components. Separate factor analyses were executed for each type of measure within each job. Several common features were evident in the results. First, the unique or specific categories for each job tended to load on different factors than the common categories. Second, the factors that emerged from the common categories tended to be fairly similar across the nine different jobs and across the three methods. Since some of the categories were not sampled in one or more of the tests for some jobs, at least some differences were inevitable.

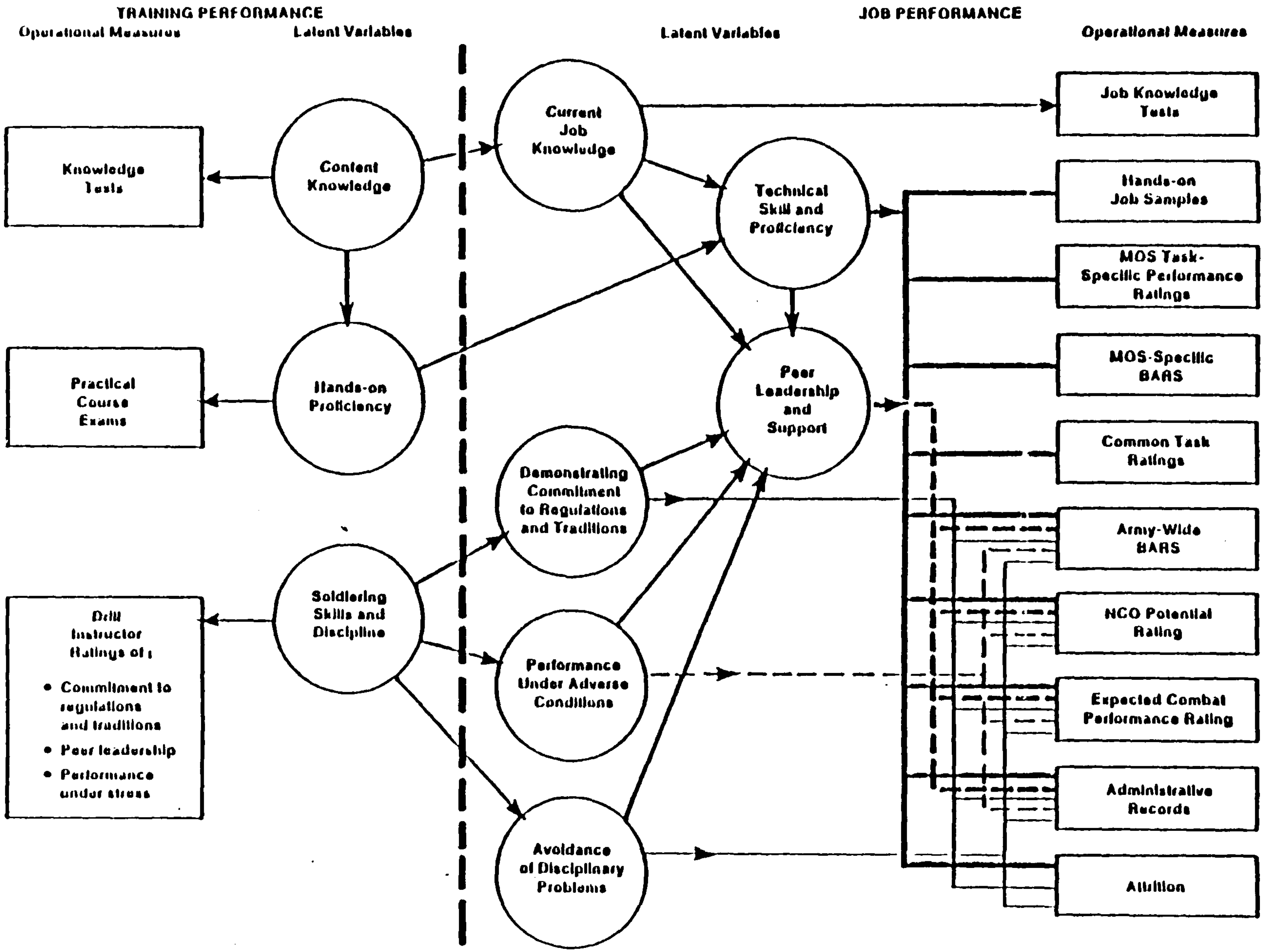


Figure 7.1. Preliminary Model of Enlisted Job Performance.

Performance Measures Common to Batch A and Batch Z MOS (Jobs)

1. Ten behaviorally anchored rating scales designed to measure factors of non-job-specific performance (e.g., *giving peer leadership and support*, maintaining equipment, self-discipline).

2. Single scale rating of overall job performance.

3. Single scale rating of NCO (noncommissioned officer) potential.

4. A 40-item summated rating scale for the assessment of expected combat performance.

5. Paper-and-pencil test of training achievement developed for each of the 19 MOS (130-210 items each).

6. Five performance indicators from administrative records, the first four obtained via self-report and the last from computerized records.

   - Total number of awards and letters of commendation.
   - Physical fitness qualification.
   - Rifle marksmanship qualification score.
   - Number of disciplinary infractions.
   - Promotion rate (in deviation units).

Performance Measures for Batch A Only

7. Job-sample (hands-on) test of MOS-specific task proficiency. Individual is tested on each of 15 major job tasks in an MOS.

8. Paper-and-pencil job knowledge tests designed to measure task-specific job knowledge. Individual is scored on 150-200 multiple choice items representing 30 major job tasks. Fifteen of the tasks were also measured hands-on.

9. Rating scale measures of specific task performance on the 15 tasks also measured with the knowledge tests and the hands-on measures.

10. MOS-specific behaviorally anchored ratings scales. From 6 to 12 BARS were developed for each MOS to represent the major factors that constituted job-specific technical and task proficiency.

Performance Measures for Batch Z Only

11. Ratings of performance on 11 representative "common" tasks. The Army specifies a series of common tasks (e.g., several first aid tasks) that everyone should be able to perform.

12. Single scale rating on performance of specific job duties.

Auxiliary Measures Included in Criterion Battery

13. Job History Questionnaire which asks for information about frequency and recency of performance of the MOS-specific tasks.

14. Army Work Environment Questionnaire - a 99-item questionnaire assessing situational/environmental characteristics, and leadership climate.

15. Measurement Method Rating obtained from all participants at the end of the final testing session.

Note: All rating measures were obtained from approximately two supervisors and three peers for each ratee.

Figure 7.2. Criterion measures used in Concurrent Validation samples.

With these exploratory empirical factor analyses used as a guide, the following set of content categories was identified:

1. Basic Soldiering Skills (field techniques, weapons, navigate, customs and laws).
2. Safety/Survival (first aid, nuclear-biological-chemical safety).
3. Communications (radio operation).
4. Vehicle Maintenance.
5. Identify Friendly/Enemy Aircraft and Vehicles.
6. Technical Skills (specific to the job).

At this point, the categories derived from the written and hands-on tests reflected an integration of expert judgment and the results of the factor analyses. (See Chapter 5 for a full description of the categories and how they were developed.)

Reduction of the Rating Variables

Army-Wide BARS. Empirical factor analyses of the behaviorally anchored Army-wide rating scales suggested three factors:

1. Effort/Leadership: including effort and competence in performing job tasks, leadership, and self-development.
2. Maintaining Personal Discipline: including self-control, integrity, and following regulations.
3. Fitness and Bearing: including physical fitness, and maintaining proper military bearing and appearance.

MOS-Specific BARS. Similar exploratory factor analyses were conducted for the job-specific BARS, and two factors within each job were identified. The first consisted of scales reflecting performance that seemed to be most central to the specific technical content of the job. The second consisted of the rating scales that seemed to reflect more tangential or less central performance components. Again the final formulation of factors was based on a combination of empirical and judgmental considerations.

Task Ratings. The reliabilities, intercorrelations, and distributional properties of the task ratings for each of the 30 tasks that were also tested with the knowledge tests were examined and found to be less reliable than either the Army-wide or the MOS-specific scales. Supervisors and peers often reported that they had never had an opportunity to observe their ratees' performance on many of the tasks, leading to a significant missing data problem. Consequently, the task ratings were dropped from the present analyses.

Combat Prediction Scale. The individual items in the combat performance prediction battery also were subjected to a principal components analysis. Two factors seemed to emerge from an analysis on the combined sample. The first factor consisted of items depicting exemplary effort, skill, or dependability under stressful conditions. The second factor consisted of items portraying failure to follow instructions and lack of discipline under stressful conditions.

Reduction of the Administrative Measures

The way in which the administrative measures were scored during Concurrent Validation was generally similar to that used in earlier testing. However, based on an examination of the intercorrelations and distributional properties of the indexes, the awards and certificates items used earlier were combined into one score, a measure of reenlistment eligibility was dropped, and a promotion rate score was developed from existing computer file information. The latter score is a deviation score in that each individual's promotion progress is compared to the mean. After these changes were made, five scores (awards/certificates, physical readiness, M16 qualification, Articles 15/flag actions, promotion rate) were identified as best serving to capture the relevant variance in the administrative indexes.

The Final Array

Based on the above exploratory analyses, the reduced array of criterion variables is shown in Table 7.1. Because MOS do differ in their task content, not all 31 variables were scored in each MOS and there was some slight variation in the number of variables used in the subsequent analyses. Table 7.2 shows the means, standard deviations, and intercorrelations among the variables for one of the nine jobs (MOS 11B, Infantryman); the summary statistics for all nine MOS are shown in Appendix D.

BUILDING THE TARGET MODEL

The next step was to build a target model of job performance that could be tested for goodness of fit within each of the nine jobs. The initial model shown in Figure 7.1 was a starting point. The correlation matrixes shown in Appendix D were each subjected to another round of empirical factor analysis to suggest possible modifications.

Several consistent results were observed in the different factor analyses. First, as expected, there was the general prominence of "method" factors, specifically one methods factor for the ratings and one methods factor for the written tests. The emergence of method factors was anticipated and was consistent with prior findings (e.g., Landy & Farr, 1980).

The second consistent result was a correspondence between the administrative measures scales and the three Army-wide rating factors. The awards and certificates scale from the administrative measures loaded together with the Army-wide effort/leadership rating factor; the Articles 15 score and the promotion rate scale loaded with the personal discipline factor (most of the variance in promotion rate was thought to be due to retarded advancement

Table 7.1

Thirty-One Basic Criterion Scores Obtained by *Aggregating* Individual Rating Scales, Job Sample Tasks, Knowledge Test Items, and Archival Records

---

1. Single scale rating of overall performance.

Three-Unit Weighted Factor Scores Obtained from the 10 Factor Analysis Army-Wide Behaviorally Anchored Rating Scales.

2. Effort and leadership factor.
3. Personal discipline factor.
4. Physical fitness and military bearing factor.

Two-Unit Weighted Factor Scores Obtained Via Factor Analysis of the Job-Specific Behaviorally Anchored Rating Scales Developed for Each Job.

5. Core responsibilities factor.
6. Peripheral responsibilities factor.

Two-Unit Weighted Factor Scores Outlined from the Expected Combat Performance Summated Rating Scale.

7. Performing well under adverse conditions factor.
8. Avoiding mistakes factor.

Archival/Administrative Performance Indicators.

9. Awards and certificates.
10. Physical readiness test score.
11. M16 qualification score.
12. Articles 15/flag actions.
13. Promotion rate deviation score.

Task Proficiency Scale Scores Obtained by Clustering Items for Hands-On Job Sample Tests (HO).

14. Core technical (MOS-specific).
15. Communications.
16. Vehicle operation and maintenance.
17. General soldiering.
18. Identifying target and threat vehicles and aircraft.
19. Safety and survival.

Job Knowledge Scale Scores Obtained by Clustering Items From Job Knowledge Tests (JK).

20. Core technical (MOS-specific).
21. Communications.
22. Vehicle operation and maintenance.
23. General soldiering.
24. Identifying target and threat vehicles and aircraft.
25. Safety and survival.

Training Knowledge Scale Scores Obtained by Clustering Items From Training School Knowledge Tests (SK).

26. Core technical (MOS-specific).
27. Communications.
28. Vehicle operation and maintenance.
29. General soldiering.
30. Identifying target and threat vehicles and aircraft.
31. Safety and survival.

Table 7.2

Job Performance Measure Summary Statistics for MOS 11B: Infantryman (N=503)

No.	Variable[a]	MN	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
				Intercorrelations[b]																								
1	Overall Rtng	4.60	.90	.	90	74	68	77	85	75	65	23	12	17	-35	36	26	14	04	35	25	11	10	33	19	18	12	14
2	Eff/Ldr Rtng	4.41	.82	90	.	74	65	80	88	80	67	24	08	13	-30	36	30	12	05	36	27	10	13	33	20	20	09	17
3	Discipl Rtng	4.50	.87	74	74	.	49	55	71	63	66	13	03	07	-39	31	16	10	03	30	22	06	08	24	13	13	05	13
4	Fitness Rtng	4.86	.89	68	65	49	.	59	66	52	45	17	27	09	-24	22	10	09	-01	10	10	-02	-04	13	06	06	01	01
5	Job-Spec Tech	32.98	4.58	77	80	55	59	.	86	75	58	23	15	17	-20	22	27	15	05	35	22	12	10	36	21	23	09	16
6	Job-Spec Oth	22.67	3.66	85	88	71	66	86	.	80	67	25	08	14	-28	32	23	10	06	35	26	12	12	33	17	22	11	17
7	Combat Exmpl	9.02	1.49	75	80	63	52	75	80	.	75	24	08	13	-31	29	28	12	07	37	25	09	16	34	22	23	09	19
8	Combat Prob	10.03	1.64	65	67	66	45	58	67	75	.	14	08	06	-33	27	20	07	-01	36	24	09	15	31	21	18	08	14
9	Awards/Certs	3.33	2.18	23	24	13	17	23	25	24	14	.	15	20	-02	04	13	06	-01	14	15	-00	13	09	09	05	04	12
10	Phys. Read	273.44	28.00	12	08	03	27	15	08	08	-08	15	.	11	02	-06	01	-07	-09	00	05	-07	-00	08	-02	-01	-04	-06
11	M16 Qualif	2.74	.57	17	13	07	09	17	14	13	06	20	11	.	01	01	13	06	-00	10	02	03	00	14	10	05	03	06
12	Article 15	.39	.85	-35	-30	-39	-24	-20	-28	-31	-33	-02	02	01	.	-45	10	-01	-06	10	-09	-08	-06	-10	-01	-09	00	-05
13	Promot Rate	.03	.68	36	36	31	22	22	32	29	27	04	-06	01	-45	.	16	07	07	19	17	12	10	18	14	12	11	17
14	HO Basic	50.50	10.06	26	30	16	10	27	23	28	20	13	01	13	-10	16	.	15	06	44	30	13	27	40	24	20	16	30
15	HO Safety	22.67	3.41	14	12	10	09	15	10	12	07	06	-07	06	-01	07	15	.	02	16	08	01	08	16	07	03	03	04
16	HO Comm	13.15	1.53	04	05	03	-01	05	06	07	-01	-01	-09	-00	-06	07	06	02	.	04	06	-01	-03	-	04	06	02	-01
17	JK Basic	50.93	9.71	35	36	30	10	35	35	37	36	14	00	10	-10	19	44	16	04	.	68	40	42	65	50	40	30	35
18	JK Safety	20.02	4.31	25	27	22	10	22	26	25	24	15	05	02	-09	17	30	08	06	68	.	23	26	47	41	32	25	20
19	JK Comm	4.37	1.47	11	10	06	-02	12	12	09	09	-00	-07	03	-06	12	13	01	-01	40	23	.	16	26	25	19	14	18
20	JK Identify	8.25	2.24	10	13	08	-04	10	12	16	15	13	-00	00	-06	10	27	08	-03	42	26	16	.	31	24	18	16	37
21	SK Basic	72.87	14.89	33	33	24	13	36	33	34	31	09	08	14	-10	18	40	16	00	65	47	26	31	.	63	60	44	43
22	SK Safety	9.51	2.12	19	20	13	06	21	17	22	21	09	-02	10	-01	14	24	07	04	50	41	25	24	63	.	45	34	26
23	SK Comm	5.68	1.67	18	20	13	06	23	22	23	18	05	-01	05	-09	12	20	03	06	40	32	19	18	60	45	.	40	31
24	SK Vehicle	.78	.42	12	09	05	01	09	11	09	08	04	-04	03	00	11	16	03	02	30	25	14	16	44	34	40	.	21
25	SK Identify	2.80	1.16	14	17	13	01	19	17	19	14	12	-06	06	-05	17	30	04	-01	35	20	18	37	43	26	31	31	.

[a]HO = Hands-On; JK = Job Knowledge; SK = School Knowledge

[b]Decimals have been omitted from the correlations.

A third observation from the empirical factor analyses was that, with the possible exception of the job-specific content factors, there was not much evidence that the factors reflecting task performance crossed measurement methods. The hands-on communication score, for example, was likely to be as correlated with the written safety score as with the written communications score. We interpreted this result as evidence against being able to separate the measurement of knowledge and the measurement of performance skill within the common task domain.

Based on these findings from the exploratory empirical analyses, we constructed a revised model to account for the correlations among our performance measures. This model included the five job performance constructs defined in Figure 7.3:

1. Core Technical Proficiency
2. General Soldiering Proficiency
3. Effort and Leadership
4. Personal Discipline
5. Physical Fitness and Military Bearing

Several minor issues remained before the model could be tested for goodness of fit within the nine Batch A jobs. One was whether the job-specific BARS were measuring job-specific technical knowledge and skill, or effort and leadership, or both. The intercorrelations among our performance factors suggested that these rating scales were measuring both of these performance constructs, though they seemed to correlate more highly with other measures of effort and leadership than with measures of job-specific technical knowledge and skill. For purposes of model fitting the MOS-specific BARS core factor was hypothesized to load on both core technical and effort/leadership.

Another issue was whether it was necessary to posit hands-on and administrative measures "method" factors to account for the intercorrelations within each of these sets of measures. The average intercorrelation among the scores within each of these sets was not particularly high. Therefore, for the sake of parsimony, we decided to try to fit a model without these two additional methods factors.

CONFIRMATORY ANALYSIS

Confirmation of the Model Within Each Job

The next step in the analysis was to conduct separate tests of goodness of fit of this target model within each of the nine jobs. This was done using the LISREL confirmatory factor analysis program (Jöreskog & Sorbom, 1981).

In conducting a confirmatory factor analysis with LISREL, it is necessary to specify the structure of three different parameter matrixes: Lambda-Y, the hypothesized factor structure matrix (a matrix of regression coefficients for predicting the observed variables from the underlying

1. <u>Core Technical Proficiency</u>
This performance construct represents the proficiency with which the soldier performs the tasks that are "central" to the MOS. The tasks represent the core of the job and they are the primary definers of the MOS. For example, the first-tour Armor Crewman starts and stops the tank engines; prepares the loader's station; loads and unloads the main gun; boresights the M60A3; engages targets with the main gun; and performs misfire procedures. This performance construct does not include the individual's willingness to perform the task or the degree to which the individual can coordinate efforts with others. It refers to how well the individual can execute the core technical tasks the job requires, given a willingness to do so.

2. <u>General Soldiering Proficiency</u>
In addition to the core technical content specific to an MOS, individuals in every MOS also are responsible for being able to perform a variety of general soldiering tasks -- for example, determines grid coordinates on military maps; puts on, wears, and removes M17 series protective mask with hood; determines a magnetic azimuth using a compass; collects/reports information - SALUTE; and recognizes and identifies friendly and threat aircraft. Performance on this construct represents overall proficiency on these general soldiering tasks. Again, it refers to how well the individual can execute general soldiering tasks, given a willingness to do so.

3. <u>Effort and Leadership</u>
This performance construct reflects the degree to which the individual exerts effort over the full range of job tasks, perseveres under adverse or dangerous conditions, and demonstrates leadership and support toward peers. That is, can the individual be counted on to carry out assigned tasks, even under adverse conditions, to exercise good judgment, and to be generally dependable and proficient? While appropriate knowledges and skills are necessary for successful performance, this construct is meant only to reflect the individual's willingness to do the job required and to be cooperative and supportive with other soldiers.

4. <u>Personal Discipline</u>
This performance construct reflects the degree to which the individual adheres to Army regulations and traditions, exercises personal self-control, demonstrates integrity in day-to-day behavior, and does not create disciplinary problems. People who rank high on this construct show a commitment to high standards of personal conduct.

5. <u>Physical Fitness and Military Bearing</u>
This performance construct represents the degree to which the individual maintains an appropriate military appearance and bearing and stays in good physical condition.

Figure 7.3. Definitions of the Job Performance Constructs.

latent constructs); Theta-Epsilon, the matrix of uniqueness or error components (and intercorrelations); and Psi, the matrix of covariances among the factors.

In these analyses, we set the diagonal elements of Psi (i.e., the factor variances) to 1.0, forcing a "standardized" solution. This meant that the off-diagonal elements in Psi would represent the correlations among and between our performance constructs and method factors. We further specified that the correlation among the two method factors and each performance construct should be zero. This effectively defined the method factor as that portion of the common variance among measures from the same method that was not predictable from (i.e., correlated with) any of the other related factor or performance construct scores.

Some problems were encountered in fitting the hypothesized model to several of the jobs. Solutions were obtained with some factor loadings greater than one and with negative uniqueness estimates for the corresponding observed variables. Also, estimates of the correlations among the performance constructs occasionally exceeded unity. These problems necessitated a certain amount of ad hoc cutting and fitting in the form of computing the squared multiple correlation (SMC) for predicting each observed variable from all of the other variables, and setting the uniqueness estimates (i.e., Theta-Epsilon diagonal) to 1.0 minus this SMC. This approach eliminated all factor loadings and correlations greater than one. In most cases, a second "iteration" was performed to adjust the initial uniqueness estimates (Theta-Epsilon) so that the diagonal of the estimated correlation matrix would be as close to 1.0 as possible.

Table 7.3 shows the final factor loading estimates (from Lambda-Y) for each job. Table 7.4 shows the uniqueness estimates (from Theta-Epsilon) and Table 7.5 shows the factor intercorrelation estimates (from Psi).

LISREL also computes a goodness-of-fit index based on a comparison of the actual correlations among the observed variables and the correlations estimated from Lambda-Y, Theta-Epsilon, and Psi. The goodness of fit is distributed as chi-square, with degrees of freedom dependent on the number of observed variables and the number of parameters estimated. The expected value of chi-square is equal to the degrees of freedom; it is a sign that the model does not fit the correlations among the observed variables.

Table 7.6 shows the value of chi-square for each job from this computation. The chi-square values should be interpreted with considerable caution because the approach we used was not purely confirmatory. The hypothesized target model was based in part on analyses of these same data; in addition, LISREL was "told" that the Theta-Epsilon (uniqueness) parameters all were fixed, and therefore did not "use up" degrees of freedom estimating these parameters; in fact, these values were estimated entirely from the data.

Table 7.3

Factor Loadings: Separate Model for Each Job

Construct/Factor[a]	MOS 11B	13B	19E	31C	63B	64C	71L	91A	95B
Core Technical									
HO Technical	--	.61	.47	.64	.51	.29	.77	.59	.32
JK Technical	--	.75	.78	.79	.74	.26	.78	.75	.32
SK Technical	--	.70	.79	.73	.82	.55	.229	.81	.43
MOS Tech Rating	--	.45	.10	.22	.25	.25	.34	.10	.13
General Soldiering									
HO Soldier	.60	.51	.46	.64	.17	.50	.60	.42	.60
HO Safety	.26	.33	.32	.31	.12	.63	.37	.48	.47
HO Communications	.05	.06	.39	.56	--	--	--	--	.80
HO Vehicle	--	--	--	.22	.17	[b]	--	--	.31
JK Soldier	.76	.52	.74	.62	.45	.48	.87	.58	.46
JK Safety	.55	.37	.75	.38	.71	.51	.72	.58	.33
JK Communications	.30	.23	.66	.38	--	--	--	--	.29
JK Vehicle	--	.17	--	.10	.41	[b]	--	--	.35
JK Identify	.46	--	.20	.28	--	.12	--	.24	.21
SK Soldier	.73	.45	.67	.39	.78	.56	.45	.44	.42
SK Safety	.47	.32	.53	.62	.57	.47	.30	.64	.32
SK Communications	.42	.26	.42	--	.41	.35	.20	--	.20
SK Vehicle	.22	.24	.05	.30	.61	[b]	.22	.47	.28
SK Identify	.46	--	.46	.13	--	--	--	--	--
Effort/Leadership									
Eff/Ldr Rating	.76	.56	.85	.64	.68	.83	.66	.76	.70
MOS Tech Ratings	.70	--	.63	.40	.41	.50	.25	.59	.52
MOS Other Rating	.77	.41	.48	.43	.54	.62	.43	.61	.56
Combat Exmplry	.80	.47	.68	.54	.57	.87	.63	.80	.77
Combat Problems	.48	.20	--	.39	.52	.53	.55	--	.56
Awards/Certificate	.32	.23	.24	.19	.28	.25	.34	.34	.22
Overall Rating	.46	.39	.33	.17	.57	.42	.65	--	.41
Discipline									
Discipline Rating	.77	.58	.73	.45	.63	.85	.74	.58	.73
Combat Problems	.29	.16	.62	.03	.05	.19	--	.82	.33
Articles 15	-.63	-.61	-.55	-.62	-.65	-.47	-.69	-.46	-.60
Promotion Rate	.74	.61	.68	.79	.63	.57	.59	.54	.54
Overall Rating	.39	.20	.53	.54	.09	.42	.06	.75	.38

(Continued)

Table 7.3 (Continued)

Factor Loadings: Separate Model for Each Job

	MOS								
Construct/Factor[a]	11B	13B	19E	31C	63B	64C	71L	91A	95B
Fitness/Bearing									
Fitness Rating	.69	.23	.84	.48	.54	.42	.50	.60	.78
Phys Readiness	.11	.90	.49	.89	.70	.53	.76	.69	.69
Ratings Method									
AW Ratings	.60	.73	.47	.70	66	54	65	66	.66
MOS Ratings	.73	.73	.60	.69	.67	49	69	54	.63
Combat Ratings	.47	.65	.55	.69	.57	.27	.55	.47	.40
Written Method									
JK Technical	--	.47	.28	.55	.59	.73	.44	.58	.57
JK Soldier	.41	.51	.33	.40	.61	.57	.11	.37	.59
JK Safety	.37	.52	.12	.63	.08	.49	.17	.76	.57
JK Communications	.34	.11	.07	.55	--	--	--	--	.52
JK Vehicle	--	--	--	.42	.62	[b]	--	.24	.21
JK Identify	-.15	.23	.50	.36	--	.05	--	.08	.23
SK Technical	--	.48	.48	.55	.46	.88	.42	27	.50
SK Soldier	.50	.66	.54	.59	.15	.51	.54	--	.54
SK Safety	.53	.55	.42	.29	.34	.48	.44	.19	.60
SK Communications	.51	.47	.46	--	.16	.24	.05	--	.42
SK Vehicle	.49	.57	.24	.48	.55	[b]	.38	.05	.42
SK Identify	.21	--	.42	.44	--	--	--	--	--
M16 Qualification	.71	.71	.71	.71	.71	.71	.71	.71	.71

[a]HO = Hands-on; JK = job knowledge; SK = school knowledge; AW = Army-wide.
[b]Vehicle content was merged into the Core Technical factor for MOS 64C.

Table 7.4

Uniqueness Estimates: Separate Model for Each Job

	MOS								
Factor Score[a]	11B	13B	19E	31C	63B	64C	71L	91A	95B
HO Technical	--	.52	.71	.48	.64	.74	.33	.57	.88
HO Soldier	.59	.66	.75	.52	.95	.74	.55	.76	.63
HO Safety	.92	.85	.75	.52	.95	.59	.79	.71	.77
HO Communications	.95	.95	.81	.62	--	--	--	--	.82
HO Vehicle	--	--	--	.03	.95	b	--	--	.90
JK Technical	--	.21	.30	.15	.12	.39	.17	.11	.53
JK Soldier	.10	.43	.22	.26	.29	.74	.31	.58	.43
JK Safety	.32	.53	.32	.31	.45	.49	.44	.15	.57
JK Communications	.56	.93	.32	.34	--	--	--	--	.64
JK Vehicle	--	--	--	.56	.32	b	--	.94	.82
JK Identify	.36	.89	.40	.51	--	.95	--	.92	.90
SK Technical	--	.27	.13	.09	.10	.14	.14	.15	.52
SK Soldier	.09	.37	.14	.48	.31	.42.	.54	.74	.46
SK Safety	.46	.59	.43	.41	.50	.55	.72	.47	.55
SK Communications	.40	.72	.35	--	.65	.82	.78	--	.67
SK Vehicle	.73	.62	.69	.55	.18	b	.73	.76	.75
SK Identify	--	.45	.10	.22	.25	.25	.34	.10	.13
Overall Rating	.13	.13	.13	.13	.13	.13	.13	.13	.18
Eff/Ldr Rating	.11	.11	.11	.11	.11	.05	.11	.11	.05
Discpln Rating	.22	.22	.22	.22	.22	.05	.22	.22	.06
Fitness Rating	.38	.38	.38	.38	.38	.05	.38	.38	.05
MOS Tech Ratings	.08	.11	.13	.14	.08	.37	.17	.12	.33
MOS Other Rating	.10	.13	.17	.19	.12	.35	.20	.18	.27
Combat Exmplry	.02	.02	.02	.02	.02	.14	.02	.02	.02
Combat Problems	.13	.13	13	.13	.13	.60	.13	.13	.40
Awards/Certif	.89	.94	.93	.95	.91	.94	.86	.85	.90
Phys Readiness	.95	.33	.67	.34	.50	.83	.46	49	.49
Articles 15	.58	.59	.68	.60	.56	.76	.51	.75	.64
Promotion Rate	.45	.60	.53	.41	.57	.64	.62	.67	.70
M16	.50	.50	.50	.50	.50	.50	.50	.50	.50

[a]HO = hands-on; JK = job knowledge; SK = school knowledge.
[b]Vehicle content was merged into the Technical factor for MOS 64C.

Table 7.5

Estimated Construct Correlations: Separate Model for Each Job

First Construct	Second Construct[a]	11B	13B	19E	31C	63B	64C	71L	91A	95B
		MOS								
Core Technical	General Sldg	--	.77	.83	.63	.58	.73	.48	.66	.70
	Effort/Lead	.67	.86	.51	.44	.50	.78	.44	.35	.46
	Discipline	.42	.13	.37	.26	.12	.69	.19	.43	.50
	Fitness	.25	.01	.03	.04	-.18	-.09	.10	-.05	-.09
	M16	.27	.00	.04	.11	.05	.05	-.09	-.17	-.10
General Soldiering	Effort/Lead	--	.89	.58	.57	.53	.44	.37	.43	.40
	Discipline	--	.29	.45	.30	.29	.29	.04	.37	.24
	Fitness	--	-.19	.05	-.05	-.03	-.14	.09	-.05	.00
	M16	--	-.06	.30	.30	.04	.11	.27	.02	.02
Effort/ Leadership	Discipline	.49	.67	.62	.55	.65	.51	.51	.59	.39
	Fitness	.57	.04	.38	-.11	.10	.23	.32	.21	.42
	M16	.38	-.13	.21	.24	-.02	.35	.22	.17	.28
Discipline	Fitness	.33	.05	.24	.24	.30	.30	.27	.19	.25
	M16	-.12	-.25	-.30	.09	-.28	-.11	.02	-.28	-.08
Fitness	M16	.52	.26	-.05	.02	.19	.22	.18	.27	.26

[a]The M16 qualification score could not be assigned to any factor on the basis of its empirical loadings in any MOS. Consequently, during the initial confirmatory analysis it was carried along as a unique variable. However, since it could not subsequently be demonstrated to possess any common variance, it was dropped from the single model confirmatory analysis and was never used in the later scoring of the five criterion factors.

Table 7.6

Goodness-of-Fit Indexes: Separate Model for Each Job

MOS		Root Mean Square Residual	Chi-Square	df	p
11B	Infantryman	.061	326.2	227	.02
13B	Cannon Crewman	.057	350.0	322	.14
19E	Armor Crewman	.065	170.0	348	.999
31C	Single Channel Radio Operator	.069	369.2	375	.58
63B	Light Wheel Vehicle Mechanic	.060	332.1	296	.07
64C	Motor Transport Operator	.058	280.1	247	.07
71L	Administrative Specialist	.067	232.6	249	.77
91A	Medical Specialist	.061	277.1	275	.45
95B	Military Police	.052	470.0	374	.001

Confirmation of an Overall Model

The results of the confirmatory procedures applied to the performance measures from each job generally supported a common structure of job performance. The procedures also yielded reasonably similar estimates of the intercorrelations among the constructs and of the loadings of the observed variables on these constructs across the nine jobs.

The final step was to determine whether the variation in some of these parameters across jobs could be attributed to sampling variation. The specific model that we explored stated that (a) the correlation among factors was invariant across jobs and (b) the loadings of all of the Army-wide measures on the performance constructs and on the rating method factor were also constant across jobs.

The proposed overall model was a relatively stringent test of a common latent structure. For example, it was quite possible that selectivity differences in the different jobs would lead to differences in the correlations between the constructs. This would tend to make it appear that the different jobs require different performance models, when in fact they do not.

The LISREL multigroups option requires that the number of observed variables be the same for each job. However, for virtually every job scores were missing on at least one of the five construct categories for at least one of the three knowledge and skill measurement methods. To handle this problem, the Theta-Epsilon error estimates for these variables were set at 1.00, and the observed correlations between these variables and all the other variables were set to zero. It was thus necessary to count the number of "observed" correlations that we generated in this manner and subtract this number from the degrees of freedom when determining the significance of the chi-square goodness-of-fit statistic.

The overall model fit very well. The root mean square residual was .047, and the chi-square was 2508.1. There were 2403 degrees of freedom after adjusting for missing variables and the use of the data in estimating uniqueness. This yields a significance level of .07, not low enough to reject the model. Tables 7.7 and 7.8 show the factor loadings and uniqueness for each job under this constrained model. Table 7.9 shows the final mapping of the criterion measures on the five performance constructs.

Obtaining Criterion Factors Scores for Individuals

To obtain an individual's score on each of the five criterion constructs, the variables composing each factor were scored and combined in the following manner.

The Core Technical Proficiency construct is composed of two major components, each of which is standardized and then added to generate the criterion score. The first component is operationally defined as the sum of

Table 7.7

Factor Loadings: Single Model Across All Jobs

	MOS								
Construct/Factor[a]	11B	13B	19E	31C	63B	64C	71L	91A	95B
Core Technical									
HO Technical	--	.59	.43	.58	.46	.27	.71	.54	.29
JK Technical	--	.71	.79	.76	.57	.72	.70	.74	.37
SK Technical	--	.66	.70	.54	.73	.55	.68	.85	.42
MOS Tech Rating	--	.21	.12	.16	.25	.01	.12	.05	-.02
General Soldiering									
HO Soldier	.52	.66	.44	.52	.16	.51	.57	.35	.58
HO Safety	.20	.44	.31	.36	.10	.49	.30	.50	.41
HO Communications	.06	.12	.37	.52	--	--	--	--	.43
HO Vehicle	--	--	--	.15	.21	b	--	--	.27
JK Soldier	.95	.50	.79	.64	.42	.69	.66	.69	.49
JK Safety	.69	.36	.75	.45	.53	.66	.57	.65	.42
JK Communications	.35	.25	.59	.51	--	--	--	--	.39
JK Vehicle	--	--	--	.28	.37	b	--	.07	.34
JK Identify	.43	.21	.34	.36	--	.12	--	.39	.18
SK Soldier	.81	.40	.67	.33	.70	.50	.42	.40	.38
SK Safety	.57	.34	.45	.40	.63	.43	.31	.62	.34
SK Communications	.51	.21	.31	--	.42	.29	.17	--	.23
SK Vehicle	.35	.22	.06	.17	.65	b	.32	.36	.21
Effort/Leadership									
Eff/Ldr Rating[b]	.76	.76	.76	.76	.76	.76	.75	.76	.76
MOS Tech Rating[b]	.59	.33	.54	.50	.45	.62	.43	.62	.62
MOS Other Rating[b]	.77	.59	.33	.45	.59	.48	.47	.58	.58
Combat Exmplry[b]	.72	.72	.72	.72	.72	.72	.72	.72	.72
Combat Problems[b]	.44	.44	44	.44	.44	.44	.44	.44	.44
Awards/Cert[b]	.26	.26	.26	.26	.26	.26	.26	.26	.26
Overall Rating[b]	.48	.48	.48	.48	.48	.48	.48	.48	.48
Discipline									
Discpln Rating[b]	.69	.69	.69	.69	.69	.69	.69	.69	.69
Combat Problems[b]	.25	.25	.25	.25	.25	.25	25	.25	.25
Articles 15[b]	-.48	-.48	-.48	-.48	-.48	-.48	-.48	-.48	-.48
Promotion Rate[b]	.52	.52	.52	.52	.52	.52	.52	.52	.52
Overall Rating[b]	.28	.28	.28	.28	.28	.28	.28	.28	.28

(Continued)

Table 7.7 (Continued)

Factor Loadings: Single Model Across All Jobs

	MOS								
Construct/Factor	11B	13B	19E	31C	63B	64C	71L	91A	95B
Fitness/Bearing									
Fitness Ratings[b]	.82	.82	.82	.82	.82	.82	.82	.82	.82
Phys Readiness[b]	.37	.37	.37	.37	.37	.37	.37	.37	.37
Ratings Method									
AW Ratings[b]	.56	.56	.56	.56	.56	.56	.56	.56	.56
MOS Ratings[b]	.61	.61	.61	.61	.61	.61	.61	.61	.61
Combat Ratings[b]	.42	.42	.42	.42	.42	.42	.42	.42	.42
Written Method									
JK Technical	--	.49	.29	.54	.71	.30	.42	.49	.49
JK Soldier	-.16	.51	.29	.40	.53	.25	.28	.60	.60
JK Safety	-.07	.49	.07	.52	.26	.28	.35	.52	.52
JK Communications	.00	.11	.19	.38	--	--	--	.41	.41
JK Vehicle	--	--	--	.19	.62	b	--	.20	.20
JK Identify	-.05	.20	.12	.17	--	.10	--	.25	.25
SK Technical	--	.54	.65	.64	.49	.71	.45	.53	.53
SK Soldier	.44	.68	.58	.61	.25	.66	.50	.60	.60
SK Safety	.34	.51	.49	.57	.18	.56	.30	.59	.59
SK Communications	.51	.46	.60	--	.20	.36	.20	50	.50
SK Vehicle	.38	.57	.17	.60	.45	b	.17	.46	.46

[a]HO = hands-on; JK = job knowledge; SK = school knowledge; AW = Army-wide.
[b]Vehicle content was merged into the Core Technical factor for MOS 64C.
These loadings were constrained to be equal across all MOS.

Table 7.8

Uniqueness Estimates: Single Model Across All Jobs

	MOS								
Factor Score[a]	11B	13B	19E	31C	63B	64C	71L	91A	95B
HO Technical	--	.62	.79	.62	.76	.91	.44	.68	.90
HO Soldier	.72	.58	.80	.70	.95	.73	.64	.87	.67
HO Safety	.95	.84	.90	.87	.95	.73	.90	.75	.81
HO Communications	.95	.95	.86	.71	--	--	--	--	.82
HO Vehicle	--	--	--	.95	.95	b	--	--	.93
JK Technical	--	.23	.28	.13	.15	.32	.28	.16	.60
JK Soldier	.10	.44	.28	.40	.48	.41	.44	.47	.40
JK Safety	.48	.56	.41	.49	.62	.44	.55	.26	.54
JK Communications	.85	.91	.57	.55	--	--	--	--	.67
JK Vehicle	--	--	--	.87	.44	b	--	.95	.85
JK Identify	.71	.90	.84	.81	--	.95	--	.64	.90
SK Technical	--	.25	.10	.24	.18	.17	.27	.19	.54
SK Soldier	.13	.37	.20	.52	.41	.31	.58	.83	.49
SK Safety	.54	62	.54	.51	.55	.51	.80	.29	.54
SK Communications	.46	.75	.48	--	.77	.78	.92	--	.70
SK Vehicle	.75	.68	.95	.61	.31	b	.86	.86	.75
Overall Rating[b]	.18	.18	.18	.18	.18	.18	.18	.18	.18
Eff/Ldr Rating[b]	.09	.09	.09	.09	.09	.09	.09	.09	.09
Discipline Rating[b]	.17	.17	.17	.17	.17	.17	.17	.17	.17
Fitness Rating[b]	.05	.05	.05	.05	.05	.05	.05	.05	.05
MOS Tech Ratings[b]	.18	.34	.22	.24	.18	.18	.18	.18	.25
MOS Other Rating[b]	.05	.24	.46	.37	.05	.05	.05	.05	.27
Combat Exmplry[b]	.26	.26	.26	.26	.26	.26	.26	.26	.26
Combat Problems[b]	.29	.29	.29	.29	.29	.29	.29	.29	.29
Awards/Cert[b]	.93	.93	.93	.93	.93	.93	.93	.93	.93
Phys Readiness[b]	.83	.83	.83	.83	.83	.83	.83	.83	.83
Articles 15[b]	.77	.77	.77	.77	.77	.77	.77	.77	.77
Promotion Rate[b]	.70	.70	.70	.70	.70	.70	.70	.70	.70

[a]HO = hands-on; JK = job knowledge; SK = school knowledge.
[b]Vehicle content was merged into the Core Technical factor for MOS 64C.
These loadings were constrained to be equal across all MOS.

Table 7.9

Mapping of Performance Factors Onto Latent Performance Constructs

	Latent Performance Constructs							
	Content Constructs					Method Constructs		M16
Criterion Measure*	Core Technical Proficiency	General Soldiering Proficiency	Effort/ Leadership	Personal Discipline	Physical Fitness/ Military Bearing	Written Know-ledge Tests	Rating Scales	M16 Qualification
AWB Effort			X				X	
AWB Discipline				X			X	
AWB Fitness					X		X	
AWB Overall			X	X			X	
MOS Technical	X		X				X	
MOS Other			X				X	
Cmbt Perform Well			X				X	
Cmbt Avoid Mistake			X	X			X	
Adm Awards/Certs			X					
Adm Phys Readiness					X			
Adm M16								X
Adm Articles 15				X				
Adm Promotion Rate				X				
HO Technical	X							
HO Communications		X						
HO Vehicles		X						
HO General Soldier		X						
HO ID Threat/Target		X						
HO Safety/Survival		X						
JK Technical	X					X		
JK Communications		X				X		
JK Vehicles		X				X		
JK General Soldier		X				X		
JK ID Threat/Target		X				X		
JK Safety/Survival		X				X		
SK Technical	X					X		
SK Communications		X				X		
SK Vehicles		X				X		
SK General Soldier		X				X		
SK ID Threat/Target		X				X		
SK Safety/Survival		X				X		

Note: Within each rating instrument, all of the factors were constrained to have an equal loading on the Rating Scales method construct. For example, the Perform Well and Avoid Mistakes factors from the Combat Performance Prediction Scale were constrained to have identical loadings on the Rating Scales method construct, but this loading did not have to be the same as the loading for the Army-Wide BARS factors, the MOS-Specific BARS factors, or the Common Task Scales factors.

*AWB = Army-wide behaviorally anchored rating scales; HO = hands-on; JK = job knowledge; SK = school knowledge.

the CVBITS scores[2] from the hands-on tests, and the second component is defined as the sum of the CVBITS scores from both the job knowledge and school knowledge tests.

The General Soldiering Proficiency score is also composed of two major components, each of which is standardized and then added to generate the criterion score. The first component is operationally defined as the sum of the CVBITS scores from the hands-on test, and the second component is defined as the sum of the CVBITS scores from both the job knowledge and school knowledge tests.

The Effort/Leadership criterion factor is composed of four major components, each of which is standardized before the four are summed. The first component corresponds to the single rating for Overall Effectiveness. The second component is composed of three subcomponents. The first is one of the three factor scores derived from the Army-wide BARS scales (i.e., the Army-wide Effort/Leadership factor) and consists of the unit-weighted sum of five different scales (Technical Skill; Effort; Leadership; Maintain Equipment; Self Development). The second and third subcomponents are the two factor scores derived from the MOS-specific BARS rating scales. (It should be noted that all rating scores used in the computation of all criterion constructs are the average of the ratings provided by supervisors and peers.) The third component is the average of the two combat rating scales. Finally, the fourth component corresponds to the administrative measure identified as Total Awards/Letters.

The Personal Discipline factor is composed of two major components, each of which is standardized before the two are added. The first component is the Personal Discipline score derived from Army-wide BARS and consists of the unit-weighted sum of three different scales (Following Regulations; Integrity; Self-Control). The second component is the sum of two administrative measures, Articles 15/Flag Actions and Promotion Rate Deviation score.

The fifth criterion factor, Physical Fitness and Military Bearing, is composed of two components; again, each is standardized before they are added to generate a criterion score. The first component is the Physical Fitness and Bearing score derived from the Army-wide BARS and consists of the unit-weighted sum of two different scales (Military Appearance; Physical Fitness). The second component corresponds to the administrative measure identified as the Physical Readiness score.

---

[2]A set of content categories derived from the hands-on and knowledge test variables, where tasks and items were assigned as follows: Communication (radio operation); Vehicle maintenance; Basic soldiering skills (field techniques, weapons, navigation, customs and law); Identify (friendly and enemy aircraft and vehicles); Technical skills (specific to the job); Safety/survival (first aid, NBC).

Criterion Residual Scores

Five residual scores were then created from the five criterion factors in the following manner. A paper-and-pencil "methods" factor score was created by first summing the two paper-and-pencil knowledge tests (job knowledge and training content knowledge scores) and then partialing out the variance due to the correlation of the total paper-and-pencil test score with all non-paper-and-pencil criterion measures (i.e., hands-on scores, rating scores, and administrative record scores). This residual was defined as the paper-and-pencil method score. This variable was in turn partialed from the Core Technical Proficiency criterion factor and from the General Soldiering Proficiency factor, creating two residual scores. A similar procedure was used to create a rating method factor score which was in turn partialed from the Effort/Leadership, Personal Discipline, and Physical Fitness/Military Bearing factors, thereby creating three more residual scores.

Criterion Intercorrelations

The five criterion factor scores, the five residual criterion scores, the single rating obtained from the overall performance rating scales, and the total score from the hands-on tests were used to generate a 12 x 12 matrix of criterion intercorrelations for each MOS in Batch A. The averages of these correlations across MOS are shown in Table 7.10.

Remember that to create the residual scores the paper-and-pencil factor was partialed from the first two criterion factors and the rating method factor was partialed from the last three criterion factors. The intercorrelations of the five criterion factors are in the upper left quadrant, the intercorrelations among the five residual scores are in the lower right quadrant, and the cross-correlations are in the upper right and lower left. Also remember that the first two factors contain items from both the knowledge tests and the hands-on tests and the last three factors all contain both ratings and administrative measures.

Some noteworthy features of this 12 x 12 matrix are the following:

- The intercorrelations of the factor pairs which confound measurement method (e.g., 1 with 2 or 3 with 4) are higher, as expected, than factor pairs which do not confound method (e.g., 1 with 3 or 2 with 4). However, they are not so high that collapsing the five factors into some smaller number would be justified. In fact, as illustrated in the next chapter, factors 1 and 2, which intercorrelate .53 on the average, yield different profiles of correlations with the selection tests.

- The correlation of the overall performance rating scales with the total hands-on test score is low (.203) but it is certainly not zero. Assuming a reliability of about .60 for each measure would yield an intercorrelation of about .34 when corrected for attenuation. Consequently, there is a substantial proportion of

Table 7.10

Mean Intercorrelations Among 12 Summary Criterion Measures for the Batch A MOS

Criterion Summary Score	CTP Raw	GSP Raw	E/L Raw	PD Raw	F/B Raw	OPF	HOT	CTP Res	GSP Res	E/L Res	PD Res	F/B Res
Core Tech Prof (raw)	1.00	.53	.28	.19	.03	.24	.74	.88	.38	.47	.23	.04
Gen Soldier Prof (raw)	.53	1.00	.27	.16	.04	.21	.72	.39	.89	.45	.19	.05
Effort/Leadership (raw)	.28	.27	1.00	.59	.46	.87	.26	.35	.33	.65	.28	.19
Personal Discipline (raw)	.19	.16	.59	1.00	.33	.65	.15	.26	.23	.44	.89	.19
Fitness/Bearing (raw)	.03	.04	.46	.33	1.00	.47	.07	.03	.04	.25	.17	.92
Overall Perf Rating	.24	.21	.87	.65	.47	1.00	.20	.31	.26	.44	.33	.19
Hands-On Total	.74	.72	.26	.15	.07	.20	1.00	.82	.79	.44	.18	.09
Core Tech Prof (resid)	.88	.39	.35	.26	.03	.31	.82	1.00	.44	.45	.25	-.01
Gen Soldier Prof (resid)	.38	.89	.33	.23	.04	.26	.79	.44	1.00	.43	.21	.01
Effort/Leadership (resid)	.47	.45	.65	.44	.25	.44	.44	.45	.43	1.00	.48	.28
Personal Discipline (resid)	.23	.19	.28	.89	.17	.33	.18	.25	.21	.48	1.00	.20
Fitness/Bearing (resid)	.04	.05	.19	.19	.92	.19	.09	-.01	.01	.28	.20	1.00

common variance between the two measures but by no means do they assess the same things. Assuming for the moment that the reliable variance in each measure is relevant to performance, a reasonable conclusion is that while performance on a standardized job sample is a significant component of performance, it is by no means all of it.

- The correlations of the residualized factor 3 (Effort/Leadership residual) with the Core Technical factor, the Core Technical residual, the General Soldiering Proficiency factor, the overall rating scale, and the hands-on total score all are about the same. Also, as compared to the correlation of the Effort/Leadership raw scores with these same variables, the correlations of the Effort/Leadership residual with the Core Technical and General Soldiering Proficiency factors go up while the correlations with Personal Discipline and Physical Fitness go down. Residualizing factor 3 (by removing the rating method factor) makes it more like a "can do" factor and less like a "will do" factor.

In general, these intercorrelations seem to behave in very lawful ways and are consistent with a multidimensional model of performance.

CONCLUDING COMMENTS

Several aspects of the final structure are noteworthy. First, in spite of some confounding of factor content with measurement method, the latent performance structure appears to be composed of very distinct components. It is reasonable to expect that the different performance constructs would be predicted by different things, so validity generalization may not exist across the performance constructs within a job. If this is so, there is a genuine question of how the performance constructs should be weighted in forming an overall appraisal of performance for use in personnel decisions.

It is tempting to infer that Effort/Leadership and Personal Discipline, particularly the latter, reflect aspects of performance that are under motivational control and consequently may be better predicted by temperament or interest measures than by measures of ability or skill. This leads us to the question of whether choices such as showing up on time, staying out of trouble, and expending extra effort under adverse conditions are a function of state or trait variables. We do have considerable data to focus on the question. It is also interesting that the residual score for Effort/Leadership becomes more like a "can do" component of performance. It may be the case that raters cannot separate "can do" from "will do" when they are asked to retrospectively aggregate an individual's task performance and evaluate it. If the degree to which an individual exhibits a characteristic level of effort and consistency of performance is not task specific, then halo might indeed be substantive variance and not error.

Given the high degree of consistency across jobs in the structure of the performance measures, it is worth asking to what extent our performance model generalizes to even wider domains of jobs. Some limitations appear likely. The "general soldiering skills" construct would almost surely be quite different outside the military; perhaps it would be replaced by a more generalized job skill construct. Similarly, it is likely that the "physical fitness and military bearing" construct also would be somewhat different for civilian occupations. The remaining constructs--Technical Skill, Effort/Leadership, and Personal Discipline--all appear to be basic components of almost any job.

In generalizing to a wider domain of jobs, it is reasonable to suppose that other latent structures would fit other populations of jobs. For example, jobs that are not organized into units and that involve a great deal of written or oral communication (e.g., sales jobs) might have a different structure. It is tempting to ask how many different performance dimension structures define different populations of jobs. However, such questions go well beyond the present finding, which is that a single structure did fit the jobs studied.

Since (a) the five-factor solution is stable across jobs sampled from this population, (b) the performance constructs seem to make sense, and (c) the constructs are based on measures carefully developed to be content valid, it seems safe to ascribe some degree of construct validity to them.

Chapter 8

BASIC CONCURRENT VALIDATION RESULTS[1]

At this point both the individual differences of new recruits and the job performance of first-tour incumbents have been modeled and reliable measures developed to assess each relevant component. As described in previous chapters, 24 scores were used to assess the predictor domain and five criterion construct scores were developed to provide a comprehensive assessment of job performance. Consequently, the basic validation data generated by the Concurrent Validation are contained in the 24 x 5 correlation matrix that can be computed for each MOS in the sample. The present chapter describes these data in some detail.

HYPOTHESIZED RELATIONSHIPS BETWEEN PREDICTOR DOMAINS AND JOB PERFORMANCE CONSTRUCTS

Rather than looking just at the entire 24 x 5 correlation matrix, we grouped the predictor scores into six domains and computed the multiple correlation of the predictor scores within each domain with each of the criterion construct scores. Figure 8.1 depicts the relationships that were expected between the predictor domains and the five job performance constructs.

From the cognitive portion of the predictor space, four ASVAB composite scores were available for the General Cognitive Ability domain, a spatial battery score was available for the Spatial Ability domain, and six computer battery scores represented the Perceptual-Psychomotor domain. It was hypothesized that these cognitive predictor composite scores would be useful for predicting scores on the two "can do" performance constructs, Core Technical Proficiency and General Soldiering Proficiency. It was hypothesized that the cognitive predictor composite scores also would be useful for predicting scores on Effort and Leadership, since this construct also contained some components of "can do" performance.

The four ABLE temperament scores, the six AVOICE vocational interests scores, and the three job reward preference scores from the JOB all were intended to serve as measures of the three domains representing the non-cognitive portion of the predictor space. It was hypothesized that these predictor composites would be most useful for predicting the "will do" job performance constructs--including Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing.

---

[1]The material in this chapter is drawn from two papers under preparation: (a) "Project A Validity Results: The Relationship Between Predictor and Criterion Domains," by Jeffrey J. McHenry, Leaetta M. Hough, Jody L. Toquam, Mary Ann Hanson, and Steven Ashworth, and (b) "Validation Analysis for New Predictors," by John P. Campbell.

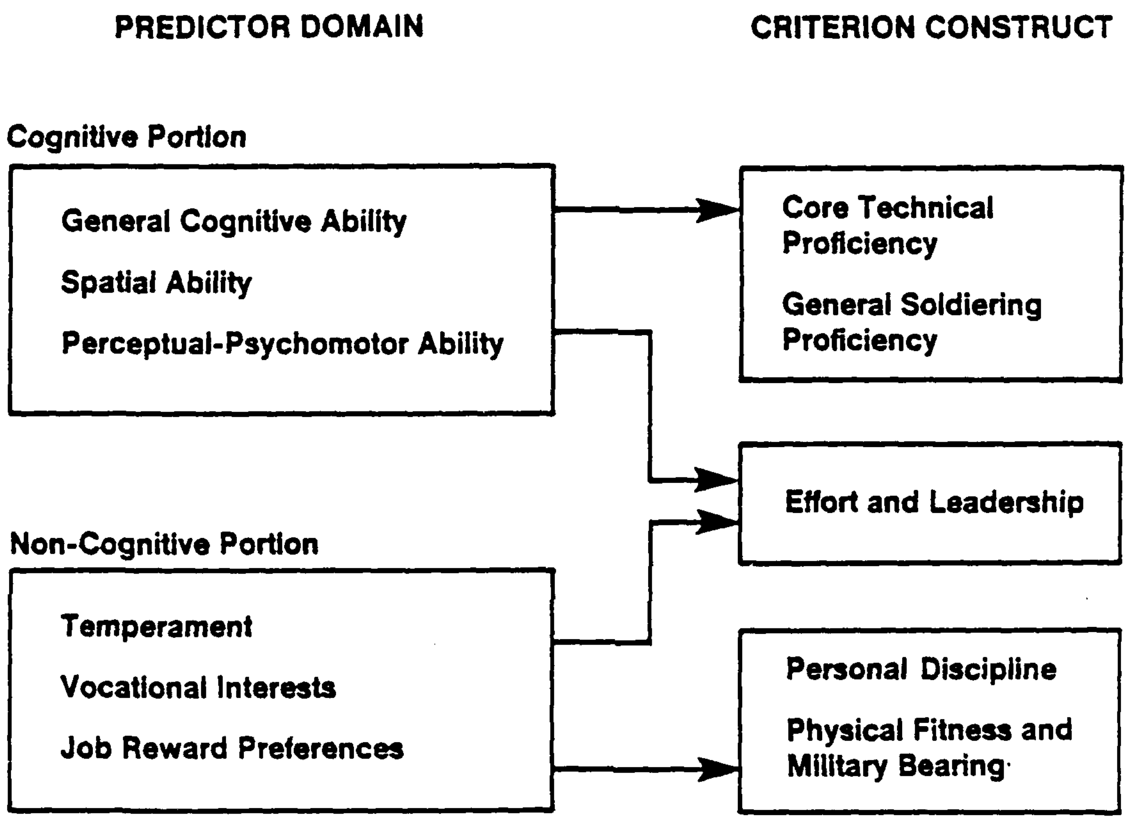


Figure 8.1. Hypothesized predictor-criterion relationships.

EMPIRICAL RELATIONSHIPS BETWEEN PREDICTOR DOMAINS AND JOB PERFORMANCE CONSTRUCTS

Statistical Estimation Procedures

To assess the relationships between predictor domains and job performance constructs, we used multiple linear regression to determine the multiple correlation $\underline{R}$ of the predictor scores within each domain with each of the five job performance constructs. This was done separately for each of the nine jobs in Batch A. Each $\underline{R}$ was corrected for range restriction and adjusted for shrinkage.

The procedure used to correct $\underline{R}$ for range restriction is one described in Lord and Novick (1968). The procedure adjusts the intercorrelations among the ASVAB subtests so that they match the intercorrelations obtained in a 1980 youth population (Mitchell & Hanser, 1984). The correlations among the predictor scores and the performance construct scores are then adjusted according to their correlations with the ASVAB subtests. This means that the correction procedure takes into account any range restriction related to the abilities measured in the ASVAB. However, it fails to consider factors that may reduce the range of predictor scores that are not related to the abilities tapped by the ASVAB.

For example, most of the soldiers in this sample enlisted in the Army between July 1983 and June 1984. They took the Project A predictor and job performance tests in the summer or fall of 1985, on average 19 months after they had reported for duty. There were some soldiers who enlisted in the Army at the same time as these soldiers, but who left the Army for various reasons before having an opportunity to be selected for the cross-validation sample. In many instances, these reasons may be unrelated to any of the abilities tapped by the ASVAB. However, several of the ABLE scales were designed to measure temperaments and traits associated with disciplinary problems. The attrition of some soldiers means that the variance of the temperament scores in our soldier sample is probably less than the variance that we would expect to obtain in an unselected sample of 18-, 19-, and 20-year-olds. Unfortunately, without data from an unselected sample, it is impossible to know the extent of this range restriction, or to correct validity coefficients for such range restriction.

Most likely, many of the validity coefficients reported in the following tables are underestimates of the true validities that would be obtained in an unselected sample. The problem is probably not very serious for the Spatial Ability composite or for the six Perceptual-Psychomotor Ability composites, which are reasonably highly correlated with scores on the ASVAB. Much of the range restriction in these composites is probably alleviated by correcting for range restriction in the ASVAB. However, the problem is more serious for the composites from the three non-cognitive predictor domains. These composites tend to be relatively uncorrelated with ASVAB scores. The validities reported for these predictor domains--and especially for the ABLE--are likely to be underestimates of the true validities.

When multiple correlation is used, the "shrinkage" problem becomes relevant. The procedure used to adjust R for shrinkage was developed by Claudy (1978). The adjustment is intended to yield an estimate of R that is equal to the expected value of the multiple correlation between the predictor scores and the criterion in the population from which the sample was drawn. The adjusted R is always lower than the observed R.

Initial Multiple Correlation Results

Given six predictor domains and five job performance constructs, 30 multiple correlations were generated for eight of the nine jobs. The Infantryman MOS was not scored on one of the performance constructs, General Soldiering Proficiency, so only 24 validity coefficients were computed for this MOS. These Rs were averaged across the nine jobs to obtain the mean validity for each predictor domain by performance construct combination.

The mean Rs for the nine MOS are reported in Table 8.1, which shows that the hypothesized predictor-criterion relationships (presented in Figure 8.1) were generally confirmed.

Table 8.1

Mean Validity[a] for the Composite Scores Within Each Predictor Domain Across Nine Army Enlisted Jobs

	Predictor Domain					
Job Performance Construct	General Cognitive Ability (K=4)[b]	Spatial Ability (K=1)	Perceptual-Psychomotor Ability (K=6)	Temperament (K=4)	Vocational Interests (K=6)	Job Reward Preferences (K=3)
Core Technical Proficiency	.63	.56	.53	.25	.35	.29
General Soldiering Proficiency	.65	.63	.57	.25	.34	.30
Effort and Leadership	.31	.25	.26	.33	.24	.19
Personal Discipline	.16	.12	.12	.32	.13	.11
Physical Fitness and Military Bearing	.20	.10	.11	.37	.12	.11

[a]Validity coefficients were corrected for range restriction and adjusted for shrinkage.
[b]K is the number of predictor scores.

The General Cognitive Ability composite, computed from the ASVAB, was the best predictor of Core Technical Proficiency (mean R = .63) and General Soldiering Proficiency (mean R = .65). These validity coefficients are extraordinarily high, especially when one considers that the ASVAB was administered to these subjects on average two years prior to the collection of job performance data. The Spatial Ability composite and the Perceptual-Psychomotor Ability composites also provided excellent prediction of Core Technical Proficiency and General Soldiering Proficiency.

The General Cognitive Ability composite also provided reasonable prediction of Effort and Leadership (mean R = .31), as hypothesized. The mean R with Effort and Leadership was only slightly lower for the composite scores from the other two cognitive domains, Spatial Ability (mean R = .25) and Perceptual-Psychomotor Ability (mean R = .26).

However, the composites from the three cognitive domains did not predict performance on Personal Discipline or Physical Fitness and Military Bearing very well. None of the six mean multiple correlations between these three predictor domains and two performance constructs exceeded .20.

The best prediction of Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing was provided by the temperament composite from the ABLE. The mean R for Effort and Leadership was .33. The ABLE score that contributed most to this correlation was Achievement Orientation. For Personal Discipline, the mean R was .32, with the ABLE Dependability score making the largest contribution to the R. Finally, the ABLE composite correlated .37 on average with Physical Fitness and Military Bearing. The key predictor of this performance construct was the ABLE Physical Condition score.

The temperament domain provided relatively lower prediction of the two "can do" performance criteria than any of the other five predictor domains. The mean R for Core Technical Proficiency was .25.

The relationships between the Vocational Interests composite and the job performance constructs were somewhat different than expected. For the interests composite, the pattern of correlations across the five job performance constructs was more like the pattern for the cognitive predictor domains than the pattern for the temperament domain. The highest mean Rs were with Core Technical Proficiency (mean R = .35). The lowest mean Rs involved prediction of Personal Discipline (mean R = .13) and Physical Fitness and Military bearing (mean R = .12). The mean validity for Effort and Leadership was .24. The pattern of correlations for the Job Reward Preference composite was similar to that for the Vocational Interests composite.

As a further test of the hypothesized predictor-criterion relationships presented in Figure 8.1, the predictor composites were grouped into two sets. The 11 General Cognitive Ability, Spatial Ability, and Perceptual-Psychomotor Ability scores were grouped into a set of cognitive composites. The 13 Temperament, Vocational Interests, and Job Reward Preferences scores were grouped into a set of non-cognitive composites. For each set the R was computed with each of the five job performance constructs within each of the nine jobs. Mean Rs from these analyses are presented in Table 8.2.

Table 8.2

Mean Validity[a] for the Cognitive, Non-Cognitive, and All Predictor Composites Across Nine Army Enlisted Jobs

Job Performance Construct	Predictor Domain: Cognitive (K=11)[b]	Non-Cognitive (K=13)	All (K=24)
Core Technical Proficiency	.65	.44	.67
General Soldiering Proficiency	.69	.44	.70
Effort and Leadership	.32	.38	.44
Personal Discipline	.17	.35	.37
Physical Fitness and Military Bearing	.23	.38	.42

[a]Validity coefficients were corrected for range restriction and adjusted for shrinkage.
[b]K is the number of predictor scores.

The pattern of correlations is very similar to that predicted in Figure 8.1. The cognitive measures provide the best prediction of Core Technical Proficiency (mean $\underline{R}$ = .65) and General Soldiering Proficiency (mean $\underline{R}$ = .69). The non-cognitive measures provide the best prediction of Personal Discipline (mean $\underline{R}$ = .35) and Physical Fitness and Military Bearing (mean $\underline{R}$ = .38). The non-cognitive scores also predict Effort and Leadership better than the cognitive scores do though the difference is not very large (mean $\underline{R}$s = .38 and .32, respectively).

Table 8.2 also shows that, when all 24 predictor scores are used to predict each performance construct, the mean $\underline{R}$s are .67 for Core Technical Proficiency, .70 for General Soldiering Proficiency, .44 for Effort and Leadership, .37 for Personal Discipline, and .42 for Physical Fitness and Military Bearing. These results indicate that for at least two of the job performance constructs--Effort and Leadership, and Physical Fitness and Military Bearing--the best prediction is obtained when both cognitive and non-cognitive predictors are used.

The one surprising result in Table 8.2 is the high correlation between the non-cognitive predictors and the two "can do" performance constructs. For both performance constructs, the mean $\underline{R}$ was .44. In fact, the non-cognitive composites predicted "can do" performance better than they predicted "will do" performance.

INCREMENTAL VALIDITY

An important question for the Army is how to improve on the validity of decisions made using the Army's current selection and classification instrument, the ASVAB. To help answer that question, the validity of the General Cognitive Ability scores (computed from the ASVAB) was compared to the validity obtained when the scores from a predictor domain were used to supplement the General Cognitive Ability composite. This was done for each performance construct within each of the nine jobs. Validities were then averaged across the nine jobs. The resulting mean validities are reported in Table 8.3.

Table 8.3

Mean Incremental Validity[a,b] for the Composite Scores Within Each Predictor Domain Across Nine Army Enlisted Jobs

	Predictor Domain					
Job Performance Construct	General Cognitive Ability (K=4)[c]	General Cognitive Ability Plus Spatial Ability (K=5)	General Cognitive Ability Plus Perceptual-Psychomotor Ability (K=10)	General Cognitive Ability Plus Temperament (K=8)	General Cognitive Ability Plus Vocational Interests (K=10)	General Cognitive Ability Plus Job Reward Preferences (K=7)
Core Technical Proficiency	.63	.65	.64	.63	.64	.63
General Soldiering Proficiency	.65	.68	.67	.66	.66	.66
Effort and Leadership	.31	.32	.32	.42	.35	.33
Personal Discipline	.16	.17	.17	.35	.19	.19
Physical Fitness and Military Bearing	.20	.22	.22	.41	.24	.22

[a]Validity coefficients were corrected for range restriction and adjusted for shrinkage.
[b]Incremental validity refers to the increase in $\underline{R}$ afforded by the new predictors above and beyond the $\underline{R}$ for the Army's current predictor battery, the ASVAB.
[c]K is the number of predictor scores.

Looking at incremental validities, Table 8.3 shows that none of the predictor domains added more than .02 to the General Cognitive Ability composite validity for predicting Core Technical Proficiency. Similarly, no predictor domain added more than .03 to the General Cognitive Ability composite validity for predicting General Soldiering Proficiency. In both instances, the predictor composite that added the greatest incremental validity was Spatial Ability. However, the four Temperament predictor scores added .11 to the validity for predicting Effort and Leadership, .19 to the validity for predicting Personal Discipline, and .21 to the validity for predicting Physical Fitness and Military Bearing.

Table 8.4 provides another means for looking at incremental validity. It shows that the seven new Project A cognitive scores (i.e., the Spatial Ability composite plus the six Perceptual-Psychomotor Ability scores) predict job performance almost as well as the four cognitive ability scores from the ASVAB. For Core Technical Proficiency and General Soldiering Proficiency, the validity of the new cognitive composites is quite high (mean $\underline{R}$ = .59 and .65, respectively). However, it is virtually identical to validities for the ASVAB. The new cognitive composites increment validity for Core Technical Proficiency by .02 and for General Soldiering Proficiency by .04. At first glance, those results may seem disappointing. However, the Army already does a very good job of predicting the "can do" components of job performance; as a practical matter, it is difficult to improve on a test with a validity of .63 or .65 for predicting job performance two years later.

Table 8.4 also shows that the 13 non-cognitive scores predict Effort and Leadership (mean $\underline{R}$ = .38), Personal Discipline (mean $\underline{R}$ = .35), and Physical Fitness and Military Bearing (mean $\underline{R}$ = .38) better than the cognitive ability composites do. When the ASVAB scores are added to the non-cognitive scores, the mean validity for Effort and Leadership increases by .05, the mean validity for Personal Discipline increases by .02, and the validity for Physical Fitness and Military Bearing increases by .03.

The results in Table 8.4 are consistent with our hypotheses that: (a) cognitive ability would be the most valid predictor of Core Technical Proficiency and General Soldiering Proficiency; (b) non-cognitive composites would be the most valid predictors of Personal Discipline and Physical Fitness and Military Bearing; and (c) both cognitive and non-cognitive predictors would be useful for predicting Effort and Leadership.

A comparison of Tables 8.3 and 8.4 shows that almost all of the incremental validity in the prediction of the three "will do" performance constructs is provided by the ABLE. When the ABLE _and_ ASVAB scores are used to predict Effort and Leadership, the mean $\underline{R}$ is .42. When the AVOICE and the JOB composites are added to the ABLE and ASVAB composites, the mean validity increases only by .01. Similarly, the AVOICE and JOB composites add only .02 to the prediction of Personal Discipline and contribute nothing to the prediction of Physical Fitness and Military Bearing.

Table 8.4

Mean Incremental Validity[a,b] for the Project A Cognitive and the Project A Non-Cognitive Predictor Composite Across Nine Army Enlisted Jobs

		Predictor Domain			
		Cognitive		Non-Cognitive	
Job Performance Construct	General Cognitive Ability (ASVAB) Composite (K=4)[c]	New Project A Cognitive Composite (K=7)	New Project A Cognitive Composite Plus ASVAB Composite (K=11)	New Project A Non-Cognitive Composite (K=13)	New Project A Non-Cognitive Composite Plus ASVAB Composite (K=17)
Core Technical Proficiency	.63	.59	.65	.44	.65
General Soldiering Proficiency	.65	.65	.69	.44	.67
Effort and Leadership	.31	.27	.32	.38	.43
Personal Discipline	.16	.13	.17	.35	.37
Physical Fitness and Military Bearing	.20	.14	.23	.38	.41

[a]Validity coefficients were corrected for range restriction and adjusted for shrinkage.
[b]Incremental validity refers to the increase in $\underline{R}$ afforded by the new predictors above and beyond the $\underline{R}$ for the Army's current predictor battery, the ASVAB.
[c]K is the number of predictor scores in the composite.

RELATIONSHIPS BETWEEN PREDICTOR DOMAINS AND "METHOD FACTORS"

The preceding chapter described written test and rating "method factors" that emerged from a structural analysis of the job performance measures. The term "method factor" is probably a misnomer since it is likely that these factors represent important components of job performance.

By definition, the written test factor reflects the variance on the paper-and-pencil tests that cannot be predicted by all the other criterion measures (ratings and hands-on). Substantively, it may reflect a soldier's comprehension of the manuals, instructions, and other materials that must be read on the job. For several of the jobs that were studied, excerpts from technical manuals and other learning aids were incorporated into the written knowledge tests.

The rating method factor is similar to what many might term "halo error." There is, however, no proof that this rating factor truly is error. It is also possible that the global impression represented by the rating factor is an important measure of soldier effectiveness. The Project A data base provides an opportunity to study the relationships between this rating factor and individual difference variables from several domains.

Table 8.5 shows the multiple correlations between the predictors within each domain and the two "method factors." The mean Rs for the written test factor are much greater than the mean Rs for the rating factor across all six predictor domains.

Table 8.5

Mean Validity[a] for the Composite Scores Within Each Predictor Domain Across Nine Army Enlisted Jobs for Written Test and Rating "Method Factor" Scores

	Predictor Domain					
Method Factor	General Cognitive Ability (K=4)[b]	Spatial Ability (K=1)	Perceptual-Psychomotor Ability (K=6)	Temperament (K=4)	Vocational Interests (K=6)	Job Reward Preferences (K=3)
Written Test	.62	.55	.54	.21	.32	.28
Rating	.15	.07	.08	.18	.09	.08

[a]Validity coefficients were corrected for range restriction and adjusted for shrinkage.
[b]K is the number of predictor scores.

The best predictor of the written test factor was the General Cognitive Ability composite (mean R = .62). Across the nine jobs the ASVAB verbal score was the most consistent predictor of the written test factor. The Spatial Ability composite and the Perceptual-Psychomotor Ability composite had mean correlations of .55 and .54, respectively. Mean correlations were lower for the three non-cognitive domains but they were not trivial, ranging from .21 for Temperament to .32 for Vocational Interests. The pattern contributes additional evidence that this factor represents a soldier's proficiency at reading job-related materials.

The best predictor of the rating factor was the Temperament composite (mean R = .18). Within the temperament domain, the most consistent predictor of the rating factor was the ABLE dependability score. The second best predictor was the General Cognitive Ability composite (mean R = .15). Mean correlations for the remaining four domains all were less than .10. This pattern of correlations suggests that the rating factor taps dependability on the job, but much more evidence would be needed to confirm this interpretation.

For Table 8.6, the predictor scores again were grouped into two sets. For the written test factor, the mean Rs across the nine jobs were .64 for the 11 cognitive scores, .40 for the 13 non-cognitive scores, and .65 across all 24 predictors. For the rating factor, the mean Rs were .16, .22, and .26, respectively.

Table 8.6

Mean Validity[a] for the Cognitive, Non-Cognitive, and All Predictor Composites Across Nine Army Enlisted Jobs for Written Test and Rating "Method Factor" Scores

	Predictor Composites		
Method Factor	Cognitive (K=11)[b]	Non-Cognitive (K=13)	All (K=24)
Written Test	.64	.40	.65
Rating	.16	.22	.26

[a]Validity coefficients were corrected for range restriction and adjusted for shrinkage.
[b]K is the number of predictor scores.

The pattern of correlations for the rating factor is similar to the pattern for the Effort and Leadership performance construct (see Table 8.2). This suggests that the rating factor obtained in this study reflects raters' global impressions of soldiers' overall competency and dependability. That is, when raters were asked to evaluate a soldier on a particular rating dimension, they considered the soldier's performance on that dimension and two other factors as well -- their general impression of how well the soldier was capable of performing the job, and their general impression of the soldier's dependability.

PREDICTOR RELATIONSHIPS WITH CRITERION RESIDUAL SCORES

Another method of studying the two method factors is to examine how the pattern of predictor-criterion relationships changes when the variance attributable to the method factors is removed from the five performance construct scores. These results are presented in Table 8.7.

The validity coefficients presented for the "raw" performance construct scores in Table 8.7 are the same as those presented in Table 8.1. To compute residual performance construct scores, the variance attributable to the written test factor was partialed from the scores for Core Technical Proficiency and General Soldiering Proficiency, and the variance attributable to the rating factor was partialed from the scores for Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing. (Written knowledge tests were not used in computing scores for Effort and Leadership, Personal Discipline, or Physical Fitness and Military Bearing. Nor were rating scales used in computing scores for Core Technical Proficiency or General Soldiering Proficiency.)

The table shows that the residual scores for Core Technical Proficiency and General Soldiering Proficiency were less predictable than the raw scores. However, the level of prediction is still substantial even when <u>all</u> variance attributable to the paper-and-pencil measurement mode is partialed out. One strong conclusion is that measurement method does not explain away the validity of ASVAB.

For Effort and Leadership, the cognitive predictor scores predicted the residual performance construct scores better than they predicted the raw performance construct scores. For example, the mean $\underline{R}$ of the General Cognitive Ability composite with the raw Effort and Leadership score was .31, while the mean $\underline{R}$ with the residual Effort and Leadership score was .46. Thus, the mean $\underline{R}$ was .15 higher for the residual score than for the raw score. The increase was .16 for the Spatial Ability composite (mean $\underline{R}$ = .41 for residual Effort and Leadership and .25 for raw Effort and Leadership) and .12 for the Perceptual-Psychomotor Ability composite (mean $\underline{R}$ = .38 and .26 for residual and raw Effort and Leadership scores, respectively).

For the ABLE composite, the results were reversed. The mean multiple correlation with the raw Effort and Leadership score was .33, while the mean $\underline{R}$ with the residual score was .31.

Table 8.7

Mean Validity[a] for the Composite Scores Within Each Predictor Domain Across Nine Army Enlisted Jobs

Job Performance Construct	Type of Score	General Cognitive Ability (K=4)[b]	Spatial Ability (K=1)	Perceptual-Psychomotor Ability (K=6)	Temperament (K=4)	Vocational Interests (K=6)	Job Reward Preferences (K=3)
		Predictor Domain					
Core Technical Proficiency	Raw	.63	.56	.53	.26	.35	.29
	Residual	.47	.37	.37	.22	.28	.21
General Soldiering Proficiency	Raw	.65	.63	.57	.25	.34	.30
	Residual	.49	.48	.41	.21	.26	.22
Effort and Leadership	Raw	.31	.25	.26	.33	.24	.19
	Residual	.46	.41	.38	.31	.32	.27
Personal Discipline	Raw	.16	.12	.12	.32	.13	.11
	Residual	.19	.15	.13	.28	.15	.10
Physical Fitness and Military Bearing	Raw	.20	.10	.11	.37	.12	.11
	Residual	.21	.11	.14	.35	.14	.10

[a]Validity coefficients were corrected for range restriction and adjusted for shrinkage.
[b]K is the number of predictor scores.

The Vocational Interests composite and the Job Reward Preferences composite actually "behaved" similarly to the Cognitive Ability composite. For both predictor domains, the mean R was greater for the residual Effort and Leadership score than for the raw Effort and Leadership score.

This pattern of correlations for Effort and Leadership suggests two interesting conclusions. First, the pattern provides additional evidence that the Vocational Interests scores are more similar to cognitive predictors than to temperament predictors.

Second, the changes in the pattern of correlations between raw and residual scores suggest that Effort and Leadership becomes more like a "can do" performance construct when the rating method factor is partialed from the raw score. The mean multiple correlations of the cognitive predictor composite with the residual Effort and Leadership score are very similar to the mean Rs of the same predictor with the two residual "can do" criterion scores. However, even after the rating factor is partialed from the raw Effort and Leadership score, the residual Effort and Leadership score continues to reflect the "will do" portion of the job performance space as suggested by its highest Rs. Thus, the residual Effort and Leadership score appears to tap both "can do" or maximal job performance and "will do" or typical job performance.

Partialing the rating factor from the Personal Discipline and the Physical Fitness and Military Bearing scores had little impact on the correlations of these scores with the predictor composites. None of the correlations for these two performance constructs changed by more than .04 when residual scores were used instead of raw scores.

RELATIVE CONTRIBUTION OF INDIVIDUAL PREDICTORS

As a first step in looking at the multiple regressions in more detail, we have portrayed the regression coefficients and zero-order validity (correlation) coefficients for individual predictor constructs in two different ways. First, Tables 8.8 and 8.9 show a stepwise solution within each of the six categories of predictor constructs (i.e., ASVAB factors, spatial factors, computer-administered factors, temperament factors, interests factors, and job reward preferences factors).

The regression equations in Table 8.8 were computed on the combined samples from the nine MOS in Batch A for each of the last four Army-wide performance factors (i.e., general soldiering/task proficiency, effort/leadership/self-development, personal discipline, and physical fitness/military bearing). The coefficients were computed on the combined samples because a series of analyses of variance had shown few predictor by MOS interactions when the dependent variable was one of the four Army-wide factors.

However, there were a number of significant predictor by MOS interactions for the Core Technical Proficiency factor. That is, the profile of regression coefficients for predicting criterion factor 1 was significantly different across MOS. The MOS by MOS stepwise regression solutions within predictor category are shown in Table 8.9.

Table 8.8

Results of Stepwise Regressions Within Each Predictor Domain for the Four Army-Wide Performance Constructs Across All Nine Batch A MOS

Predictor Construct	Criterion Construct: General Soldiering (raw score)	Effort and Leadership (resid score)	Effort and Leadership (raw score)	Personal Discipline (raw score)	Phys Fitness/ Mil Bearing (raw score)
ASVAB Factors					
Verbal	.10	.03	-.07	-.03	-.11
Quantitative	.20	.08	.03	.07	.03
Technical	.26	.21	.21	.06	-.05
Speed	.03	.07	.09	.04	.10
ADJ. UNCORR R	.461	.280	.206	.106	.161
Spatial					
Overall Spatial	.47	.25	.14	.07	-.05
UNCORRECTED R	.466	.253	.142	.068	.047
Computer					
Complex Perc Speed	-.09	-.06	-.07	--	--
Complex Perc Accy	.19	.07	.09	.05	--
Number Speed/Accy	-.14	-.06	-.09	-.03	--
Psychomotor	-.19	-.08	-.10	--	--
Simp Reaction Accy	.04	--	--	--	-.06
Simp Reaction Speed	--	--	--	--	-.07
ADJ. UNCORR R	.363	.149	.208	.032	.071
Temperament					
Adjustment	.09	.04	.03	.03	--
Dependability	.04	--	.06	.30	.12
Surgency	.04	.23	.25	--	.12
Phys Condition	-.06	--	--	-.06	.24
ADJ. UNCORR R	.129	.255	.303	.303	.356
Interests					
Combat	.24	.20	.17	--	.04
Machines	--	--	--	-.04	-.06
Audiovisual	--	--	-.04	--	--
Technical	--	.06	.08	.09	.14
Food Service	-.10	-.16	-.12	.06	-.05
Protective Svc	-.06	--	--	-.09	--
ADJ. UNCORR R	.229	.235	.199	.078	.119
Job Values					
Security	--	.03	.05	.05	.10
Autonomy	.05	.07	.03	-.06	-.05
Routine	-.11	-.12	-.09	-.03	.02
ADJ. UNCORR R	.123	.150	.112	.063	.097

Table 8.9

Results of Stepwise Regressions Within Each Predictor Domain for MOS-Specific Core Technical Proficiency for Each of the Nine Batch A MOS

	MOS								
Predictor Construct	11B	13B	19E	31C	63B	64C	71L	91A	95B
ASVAB Factors									
Verbal	.20	--	.13	.19	--	--	.16	.25	.11
Quantitative	.14	.09	.15	.14	--	.14	.38	.12	.16
Technical	.23	.23	.27	.23	.55	.34	.11	.19	.11
Speed	.10	--	--	.11	--	--	.08	.17	.09
ADJ, UNCORR R	.503	.254	.452	.427	.538	.413	.441	.456	.282
Spatial									
Overall Spatial	.48	.33	.43	.32	.41	.37	.41	.38	.28
UNCORRECTED R	.475	.334	.432	.315	.412	.366	.411	.380	.275
Computer									
Complex Perc Speed	-.25	-.10	--	--	-.08	-.14	--	--	--
Complex Perc Accy	.29	.11	.16	.13	--	.19	.27	.09	.13
Number Speed/Accy	-.11	-.11	-.20	-.25	-.08	-.07	-.22	-.20	-.19
Psychomotor	-.13	-.17	-.11	-.09	-.20	-.10	--	-.15	-.09
Simp Reaction Accy	--	--	.12	--	.08	.07	--	.08	--
Simp Reaction Speed	--	--	--	--	--	--	--	--	--
ADJ, UNCORR R	.406	.257	.343	.253	.242	.269	.325	.261	.228
Temperament									
Adjustment	--	.12	.14	--	.10	--	--	.10	.08
Dependability	--	--	.08	.10		--	.10	.19	.12
Surgency	.19	--	--	--	.09	--	.14	--	--
Phys Condition	--	--	-.13	--	-.12	--	-.10	-.15	--
ADJ UNCORR R	.143	.000	.129	.000	.119	.000	.176	.211	.114
Interests									
Combat	.25	.25	.26	--	.11	.09	.12	.18	--
Machines	--	.10	--	.13	.38	.09	-.23	--	--
Audiovisual	--	--	--	--	-.11	--	--	--	-.08
Technical	.08	--	--	.10	--	--	.19	--	--
Food Service	-.22	-.16	-.11	--	-.10	-.12	-.07	--	-.06
Protective Svc	-.11	-.10	--	--	-.14	--	--	--	--
ADJ, UNCORR R	.276	.255	.218	.000	.441	.135	.160	.039	.000
Job Values									
Security	--	--	--	--	--	--	--	.14	--
Autonomy	.08	.17	--	--	.14	.11	--	--	--
Routine	-.15	-.14	-.21	--	-.10	-.07	-.12	--	-.08
ADJ, UNCORR R	.141	.201	.166	.000	.133	.080	.038	.058	.000

For the four Army-wide components (Table 8.8), some comparisons of interest are the following:

- Among ASVAB scores the quantitative and technical scores contribute the most to the prediction of General Soldiering Proficiency. The verbal score plays a more prominent role in the prediction of the Core Technical performance factor (i.e., as shown in Table 8.9).

- While ASVAB does not contribute much to the prediction of performance factors 4 and 5, the ASVAB technical score does make a relatively large contribution to the prediction of factor 3, the Effort/Leadership factor.

- The differential contributions of the temperament (ABLE) scores to prediction of performance factors 3, 4, and 5 are clear, significant, and pronounced. The profiles look like they should.

- The combat interests score was the most predictive interest score among the scores generated from the AVOICE.

For the MOS by MOS stepwise regression coefficient profiles used to predict the core technical factor (i.e., Table 8.9), the greatest differential is within the ASVAB and the AVOICE, and to a lesser extent within the spatial and computerized tests.

To look at the coefficients in another way, the same procedures that produced Tables 8.8 and 8.9 were used to carry out stepwise regressions when all 24 predictor scores were used to predict each performance factor. Again the analyses for the four Army-wide criterion factors were carried out on a combined sample while the analyses against the core technical factor were done MOS by MOS. The results are shown in Tables 8.10 and 8.11.

Again the differential patterns appear across the four Army-wide performance factors and across MOS for the core technical factor. However, a surprise was the strong role played by the spatial factor and the combat interest factor in predicting the technical performance factor in the combat specialties.

To round out this initial picture of the contributions of individual predictor factors, the zero-order correlations (validity coefficients) corresponding to the regression coefficients in Tables 8.10 and 8.11 are shown in Tables 8.12 and 8.13.

Table 8.10

Results of Stepwise Regressions for the Four Army-Wide Performance Constructs Across All Nine Batch A MOS

Predictor Construct	General Soldiering (raw score)	Effort and Leadership (resid score)	Effort and Leadership (raw score)	Personal Discipline (raw score)	Phys Fitness/ Mil Bearing (raw score)
	Criterion Construct				
ASVAB Factors					
Verbal	.09	.03	-.06	--	-.10
Quantitative	.09	.04	--	.05	--
Technical	.12	.11	.15	.07	-.03
Speed	--	.04	.06	.03	.08
Spatial					
Overall Spatial	.25	.13	--	--	--
Computer					
Complex Perc Speed	--	--	-.05	--	--
Complex Perc Accy	.08	--	.04	--	--
Number Speed/Accy	-.02	--	--	.03	--
Psychomotor	-.04	--	-.02	--	--
Simp Reaction Accy	--	--	--	--	-.04
Simp Reaction Speed	-.03	--	--	--	-.05
Temperament					
Adjustment	--	--	--	--	--
Dependability	.11	.06	.11	.30	.09
Surgency	-.04	.15	.20	.03	.14
Phys Condition	--	.03	--	-.05	.22
Interests					
Combat	.13	.11	.10	--	.04
Machines	--	--	--	--	-.05
Audiovisual	--	-.02	-.04	-.03	.04
Technical	--	--	--	--	--
Food Service	-.04	-.08	-.06	-.04	--
Protective Svc	--	.03	--	-.03	-.05
Job Values					
Security	--	--	--	--	--
Autonomy	--	--	--	-.05	-.04
Routine	-.03	-.04	-.03	--	--
ADJ. UNCORR R	.540	.392	.366	.317	.385

Table 8.11

Results of Stepwise Regressions for MOS-Specific Core Technical Proficiency for Each of the Nine Batch A MOS

	MOS								
Predictor Construct	11B	13B	19E	31C	63B	64C	71L	91A	95B
ASVAB Factors									
Verbal	.17	--	.10	.21	--	--	.08	.26	.13
Quantitative	.09	--	--	.30	--	--	.27	--	--
Technical	.10	--	.16	--	.35	.30	-.13	.12	--
Speed	--	--	--	--	--	-.07	--	.13	--
Spatial									
Overall Spatial	.20	.25	.19	--	.14	.16	.25	.23	.22
Computer									
Complex Perc Speed	.18	--	--	--	--	-.12	--	--	--
Complex Perc Accy	.13	--	.09	-.10	--	.14	.15	--	.09
Number Speed/Accy	--	--	-.09	--	--	--	--	--	-.11
Psychomotor	--	--	--	--	--	--	--	--	--
Simp Reaction Accy	--	--	.07	--	--	--	--	--	--
Simp Reaction Speed	--	-.10	--	--	-.11	--	--	--	--
Temperament									
Adjustment	-.08	--	--	-.09	--	--	--	--	--
Dependability	.12	--	.10	.15	.13	.07	.11	.22	.12
Surgency	--	--	--	--	--	--	--	--	--
Phys Condition	--	--	-.09	--	-.06	--	--	-.13	--
Interests									
Combat	.15	.21	.17	--	--	--	-.16	.16	--
Machines	--	--	--	.21	.32	--	--	--	--
Audiovisual	--	--	--	--	-.14	--	--	-.09	-.13
Technical	--	--	--	--	--	--	.12	--	--
Food Service	-.07	--	--	--	--	--	--	--	--
Protective Svc	--	-.08	--	--	-.08	--	--	--	--
Job Preferences									
Security	--	--	--	--	--	.09	--	.12	.09
Autonomy	--	.09	--	-.11	--	--	--	--	--
Routine	-.06	-.11	--	--	--	--	--	.07	--
ADJ. UNCORR *R*	.560	.305	.464	.352	.591	.401	.481	.507	.294

Table 8.12

Correlations Between the Predictor Constructs and the Army-Wide Criterion Constructs Combined Across Batch A MOS[a]

	Criterion Construct				
Predictor Construct	General Soldiering (raw score)	Effort and Leadership (resid score)	Effort and Leadership (raw score)	Personal Discipline (raw score)	Phys Fitness/ Mil Bearing (raw score)
ASVAB Factors					
Technical	.55	.39	.28	.12	-.08
Verbal	.52	.35	.20	.10	-.07
Quantitative	.54	.36	.23	.14	-.01
Speed	.37	.29	.21	.11	.07
Cognitive Constructs					
Overall Spatial	.59	.38	.24	.11	-.03
Computer Constructs					
Complex Perc Speed	-.21	-.17	-.13	-.03	-.04
Complex Perc Accy	.30	.18	.12	.08	-.01
Number Speed/Accy	-.44	-.31	-.21	-.09	-.01
Psychomotor	-.40	-.27	-.20	-.04	-.01
Simp Reaction Accy	.18	.09	.05	.05	-.05
Simp Reaction Speed	-.19	-.13	-.08	-.01	-.06
ABLE Constructs					
Adjustment	.18	.22	.23	.13	.17
Physical Condition	-.03	.09	.10	-.02	.30
Dependability	.09	.15	.21	.30	.22
Surgency	.16	.30	.33	.20	.27
AVOICE Constructs					
Audiovisual Arts	.02	.02	.01	.00	.07
Combat Related	.23	.22	.19	.00	.03
Food Service	-.12	-.14	-.11	-.06	.00
Structural/Machines	.06	.06	.06	-.05	-.01
Protective Services	-.04	.03	.04	-.04	.02
Skilled Technical	.04	.07	.06	.05	.11
Job Constructs					
Autonomy	.13	.15	.09	-.02	-.02
Routine	-.21	-.20	-.15	-.06	-.04
Job Security	.09	.11	.10	.05	.09

[a]Corrected for range restriction.

Table 8.13

Correlations Between the Predictor Constructs and Core Technical Proficiency[a]

	MOS								
Predictor Construct	11B	13B	19E	31C	63B	64C	71L	91A	95B
ASVAB Factors									
Technical	.60	.36	.56	.59	.69	.55	.37	.61	.51
Verbal	.63	.33	.49	.67	.50	.44	.56	.71	.59
Quantitative	.60	.32	.49	.67	.45	.46	.63	.64	.59
Speed	.48	.25	.28	.57	.29	.27	.52	.56	.47
Cognitive Construct									
Overall Spatial	.63	.41	.55	.58	.56	.51	.57	.64	.56
Computer Constructs									
Complex Perc Speed	-.33	-.15	-.17	-.25	-.24	-.25	-.11	-.28	-.20
Complex Perc Accy	.35	.24	.32	.22	.16	.28	.40	.25	.26
Number Speed/Accy	-.48	-.30	-.42	-.62	-.37	-.38	-.50	-.57	-.53
Psychomotor	-.43	-.30	-.36	-.34	-.36	-.34	-.26	-.44	-.32
Simp Reaction Accy	.17	.11	.26	.17	.14	.19	.27	.16	.20
Simp Reaction Speed	-.17	-.19	-.15	-.10	-.23	-.19	-.11	-.21	-.23
ABLE Constructs									
Adjustment	.26	.13	.18	.06	.21	.07	.20	.12	.27
Physical Condition	.06	-.04	-.09	-.18	-.13	-.07	-.12	-.09	-.13
Dependability	.16	.01	.09	.04	.00	.01	.21	.18	.24
Surgency	.31	.06	.16	.14	.20	.09	.27	.22	.25
AVOICE Constructs									
Audiovisual Arts	.04	-.05	-.01	.20	-.14	.00	.19	.13	-.14
Combat Related	.23	.21	.31	.08	.31	.24	.02	.22	.03
Food Service	-.30	-.14	-.14	.01	-.20	-.14	-.03	-.09	-.19
Structural/Machines	-.12	.09	.06	.05	.41	.16	-.19	.01	-.19
Protective Svc	-.05	-.08	-.04	-.01	-.10	-.05	.01	-.13	-.16
Skilled Technical	.07	-.03	.09	.12	-.08	.00	.17	.00	-.03
Job Preferences									
Autonomy	.21	.22	.09	.22	.25	.21	.21	.23	.09
Routine	-.27	-.18	-.27	-.19	-.21	-.20	.19	.22	-.30
Job Security	.14	.13	.05	-.02	.06	.14	.20	.18	-.01

[a]Corrected for range restriction.

CONCLUDING COMMENTS

The pattern of predictor-criterion relationships presented in this chapter was consistent with the pattern that was expected. Cognitive predictors provided excellent prediction of Core Technical Proficiency and General Soldiering Proficiency. Across nine very different jobs, the mean R for the complete set of 11 cognitive composite scores was .65 and .69, respectively. Clearly cognitive predictors provide excellent prediction of job proficiency for Army enlistees. Non-cognitive predictors--in particular, temperament/biographical scores--are good predictors of Personal Discipline and Physical Fitness/Military Bearing. The best prediction of Effort and Leadership was obtained when both cognitive and non-cognitive predictors were used. In sum, the predictor-criterion relationships enhanced understanding of both the predictor space and the job performance space.

Given these initial results, the following conclusions seem warranted:

- Total job performance is multidimensional and validity evidence must be interpreted with this in mind.
- Rating scales, as a method for assessing performance, most likely reflect both the skill with which an individual can perform job tasks and the consistency and willingness with which he or she does it over some period of time.
- ASVAB does predict the "can do" aspects of performance. Its validity is not limited to the prediction of training success. In fact, it predicts job performance just as well as it predicts training performance.
- The major components of performance have very different profiles of regression/validity coefficients across the different predictor domains in the Project A Trial Battery.
- Temperament/biographical data prediction scales are valid predictors of the "will do" component of performance and in fact yield differential predictions of the three "will do" components. The Project A data suggest considerable construct validity for these scales.
- The primary source of differential prediction across jobs is that part of performance which is specific to the unique task content of the job (i.e., performance factor 1).

On the criterion side, the pattern of predictor-criterion correlations added to our confidence in the construct validity of the job performance scores. The pattern of correlations also enhanced understanding of the Effort and Leadership construct, the written test and rating method factors, and the relationship between raw and residual performance construct scores.

Continuing the data analysis, our future validity analyses will be concerned with:

- o More precise estimates of validity generalization across jobs as a function of criterion content and predictor battery composition.
- o Estimation of differential prediction across race and gender groups as a function of criterion content and predictor battery composition.
- o Estimation of overall selection validity (against a criterion composite) as a function of criterion components weights and predictor battery composition.
- o Estimation of classification efficiency.

Many of these analyses will require the generation of composite criterion scores. A single criterion score is also a requirement for calculating the predictive relationships needed for the Project B algorithms. Work related to the development of performance composite scores is already under way.

Chapter 9

SUMMARY AND FUTURE PLANS

This report has described the procedure and results of the Concurrent Validation phase of Project A. During the preceding development phase of the project, a 4-hour battery of new selection/classification tests had been developed to enable the Army to systematically sample the most relevant applicant characteristics not presently covered by ASVAB. Also during the development phase, a 12-hour training and job performance assessment procedure had been constructed so as to provide multiple measures of every major component of each job in a sample of entry-level MOS.

The sample MOS for the project were representatively selected from the population of entry-level MOS. Consequently, for (a) jobs, (b) performance components, and (c) selection/classification measures, a population had been defined and then sampled comprehensively. This makes the results of the Concurrent Validation extremely important for guiding future selection/classification practices in the Army. No other organization in the world (public or private) has such an extensive, carefully developed, and generalizable body of information on which to base personnel practices. It can be used for many years to come.

CONCURRENT VALIDATION SUMMARY

More specifically, what has this phase of Project A provided, both to the organization (the Army) and to the science (industrial and organizational psychology)? We think that so far it has achieved at least the following:

1. Much has been learned about the nature of performance in entry-level-skilled jobs. We now have a much clearer idea of what major factors constitute performance and how they can be measured. The "criterion problem" is better understood.

2. As a by-product of the analyses involving ASVAB, we have a much better, clearer idea of its factor structure and of what the factors are measuring.

3. The Concurrent Validation data support the assertion that supervisor ratings of subordinate performance can have considerable validity if a careful procedure is followed. However, the data also support the conclusion that supervisors assess, at the same time, both the technical performance of individuals and their general dependability/motivation.

4. The question of whether ASVAB does or does not predict job performance (and not just training achievement) has been definitively answered, in the affirmative. Although the subtests on the ASVAB are somewhat narrow in scope, the abilities that are measured are measured very well. The Army and the Department of Defense should now be much better able to evaluate results from these tests and use them as a basis for action.

5. Analysis of Skill Qualification Test data in relation to Project A criteria has shown that the SQT does have considerable validity as an indicator of soldier technical knowledge and proficiency.

6. The Project A job/task analysis procedures worked well and can be used by the Army in the future to develop training curricula, SQT content, and field exercises.

7. Within the constraints of a concurrent design, the CV results demonstrate that the Trial Battery yields substantial gains in selection validity. The non-cognitive measures in particular may be very valuable if they are implemented.

   It is also true that the full value of the Trial Battery can be realized only if Project A is able to carry out all the planned steps.

   The Longitudinal Validation data are needed (a) to assess responsiveness of the non-cognitive measures to experience, (b) to determine how performance in training forecasts later successes and failures on the job, (c) to determine how much the Trial Battery adds to the prediction of attrition, and (d) to determine what tests are the best predictors of success in the second tour.

   Also, a great deal of additional analysis is necessary to determine the optimal combination of ASVAB subtests and new predictors for the multiple purposes of selection accuracy, classification utility, and attrition reduction. Much time and effort has already been devoted to data analysis; however, if the Army is to realize its full gains from the available data base, there is much more to be done.

8 The CV results do in fact indicate where the greatest gains in classification can be achieved. That is, it is clear from the CV results that the gains in classification efficiency will come from the optimal use of the Core Technical prediction equations. Again, using them in an optimal way will be dependent on additional analyses.

9. Because the Trial Battery represents a carefully selected sample of measures from a population of information, it is entirely possible to play "what if" games with different selection battery combinations. For example, if the Army wanted to devote 4 hours to selection/classification testing and within that constraint wanted to maximize the accuracy of predicting who will become discipline problems, we can model such a goal. Further, we can specify what the organization would have to trade off, in terms of aggregate performance (if anything), for increased accuracy in detecting future discipline problems. Answers to a number of such "what if" questions are now possible because we chose to sample selection information from a population of information, and not to tie test development to specific jobs.

10. Several of the criterion measures could profitably be put to operational use. For example, the procedures used for developing hands-on and knowledge tests could be used to improve the procedures for constructing SQTs. More directly, perhaps, the MOS BARS dimensions and critical incidents could be used by NCOs to provide a performance feedback and job coaching procedure for their enlisted personnel. In fact, many of the NCOs participating in the Concurrent Validation wanted to use the MOS BARS scales in this way.

NEXT STEPS

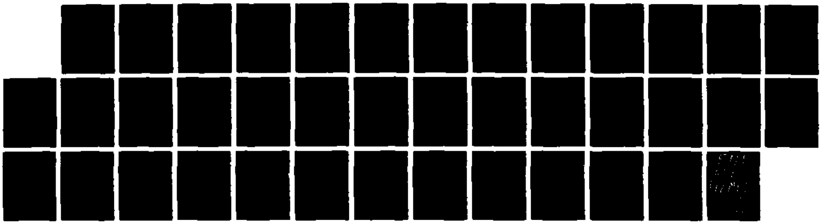
During the next fiscal year, if all goes well, Project A will pursue four principal activities:

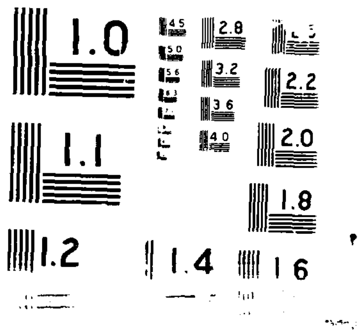
- Administering the experimental predictor battery to 45,000 new recruits with the goal of preserving a sample of 400-600 job incumbents in each of 21 MOS for Longitudinal Validation.
- Completing the job analyses and beginning the criterion development for measurement of second-tour performance in each Batch A MOS.
- Using the Concurrent Validation results to determine the extent of differential prediction across gender and racial groups, the degree of differential prediction across jobs, and the degree of classification efficiency that seems to be possible for enlisted MOS.
- Measuring the utility of performance at different levels in Army entry-level jobs.

We cannot say too often that, if the project can be completed, the Army will have a vastly enhanced personnel management capability. For example, it will be possible to determine directly how adverse impact can be minimized and prediction of performance maximized with least cost to the Army. No other organization has the capability for doing that. The same is true for a number of other critical considerations in the management of personnel. Project A is indeed unique.

AD-A198 856 IMPROVING THE SELECTION CLASSIFICATION AND UTILIZATION OF ARMY ENLISTED P..(U) HUMAN RESOURCES RESEARCH ORGANIZATION ALEXANDRIA VA J P CAMPBELL MAY 88 3/3

UNCLASSIFIED HUMRRO-IR-PRD-87-10 ARI-TR-792 F/G 5/9 NL




1.0
1.1
1.2
1.4
1.6
1.8
2.0
2.2
2.8
3.2
3.6
4.0
4.5
5.0
5.6

REFERENCES

Campbell, C. H. (In preparation). Developing basic criterion scores for hands-on tests, job knowledge tests, and task rating scales. (HumRRO IR-PRD-87-15) (ARI Technical Report in preparation).

Campbell, C. H., Campbell, R. C., Rumsey, M. G., & Edwards, D. C. (1986). Development and field test of task-based MOS-specific criterion measures (ARI Technical Report 717). (AD A182 645)

Campbell, J. P. (1986). Project A: When the textbook goes operational. Paper presented at the Annual Convention of the American Psychological Association, Washington.

Campbell, J. P. (Ed.). (1987). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1985 fiscal year (ARI Technical Report 746). (AD A193 343)

Campbell, J. P. (In preparation). Validation analysis for new predictors. Paper prepared for Data Analysis Workshop of the Committee on Performance of Military Personnel, of the National Academy of Sciences/ National Research Council Commission on Behavioral and Social Sciences and Education, December 1986.

Campbell, J. P., & Harris, J. H. (1985). Criterion reduction and combination via a participation decision-making panel. Paper presented at the Annual Meeting of the American Psychological Association, Los Angeles.

Claudy, J. G. (1978). Multiple regression and validity estimation in one sample. Applied Psychological Measurement, 2, 4, 595-601.

Davis, R. H., Davis, G. A., Joyner, J. N., & de Vera, M. V. (1987). Development and field test of job-relevant knowledge tests for selected MOS (ARI Technical Report 757). (AD A192 211)

Dunnette, M. D. (1963). A modified model for selection research. Journal of Applied Psychology, 47, 317-323.

Eaton, N. K., & Goer, M. H. (Eds.). (1983). Improving the selection, classification, and utilization of Army enlisted personnel: Technical appendix to the annual report (ARI Research Note 83-37). (AD A137 117)

Eaton, N. K., Goer, M. H., Harris, J. H., & Zook, L. M. (Eds.). (1984). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1984 fiscal year (ARI Technical Report 660). (AD A178 944)

Hoffman, R. G. (1986). Post differences in hands-on tasks. Paper presented at Annual Conference of the Military Testing Association, Mystic, CT.

Hough, L. M. (Ed.). (1988). Literature review: Utility of temperament, biodata, and interest assessment for predicting job performance (ARI Research Note 88-02). (AD A192 109)

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1983). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report (ARI Research Report 1347). (AD A141 807)

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1984). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report synopsis, 1984 fiscal year (ARI Research Report 1393). (AD A173 824)

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1984). Improving the selection, classification, and utilization of Army enlisted personnel: Appendices to annual report, 1984 fiscal year (ARI Research Note 85-14).

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (1987). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1985 fiscal year - Supplement to ARI Technical Report 746 (ARI Research Note 87-54). (AD A188 267)

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, & Army Research Institute. (In preparation). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1986 fiscal year - Supplement to ARI Technical Report (ARI Research Note in preparation).

Jöreskog, K. C., & Sorbom, D. (1981). LISREL VI: Analysis of linear squares methods. Uppsala, Sweden: University of Uppsala.

Kuhn, D. B., Schultz, S. R., & Park, M. V. Developing the basic criterion scores for job-relevant knowledge tests (HumRRO report in preparation).

Landy, F. J., & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87, 72-107.

Lord, P., & Novick, M. (1968). Statistical theory of mental test scores. Reading, MA: Addison-Wesley Publishing Company.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (In preparation). Project A validity results: The relationship between predictor and criterion domains. Paper prepared for the Annual Conference of the Society of Industrial and Organizational Psychology, Atlanta, April 1987.

McHenry, J. J., & Rose, S. R. (1988). Literature review: Validity and potential usefulness of psychomotor ability tests for personnel selection and classification (ARI Research Note 88-13). (AD A193 558)

Mitchell, K. J., & Hanser, L. M. (In preparation). The 1980 youth population norms: Enlisted and occupational classification standards in the Army (ARI Technical Report in preparation). Alexandria, VA: U.S. Army Research Institute.

Peterson, N. G. (Ed.). (1987). Development and field test of the Trial Battery for Project A (ARI Technical Report 739). (AD A184 575)

Peterson, N. G., Hough, L. M., Dunnette, M. D., Rosse, R. A., Houston, J. S., Toquam, J. L., & Wing, H. (In preparation). Identification of predictor constructs and development of new selection/classification tests. Paper prepared for the Annual Conference of the Society of Industrial and Organizational Psychology, Atlanta, April 1987.

Pulakos, E. D., & Borman, W. C. (1986a). Development and field test of Army-wide rating scales and the rater orientation and training program (ARI Technical Report 716). (AD B112 857)

Pulakos, E. D., & Borman, W. C. (1986b). Developing the basic criterion scores for Army-wide and MOS-specific ratings (ARI Technical Report in preparation).

Riegelhaupt, B. J., Harris, C. D., & Sadacca, R. (1987). The development of administrative measures as indicators of soldier effectiveness (ARI Technical Report 754). (AD A191 232)

Schmidt, F. L., & Kaplan, L. B. (1971). Composite vs. multiple criteria: A review and resolution of the controversy. Personnel Psychology, 24, 419-434.

Toquam, J. L., Corpe, V. A., Dunnette, M. D., & Keyes, M. A. (In preparation). Literature review: Cognitive abilities--theory, history, and validity (ARI Research Note in preparation).

Toquam, J. L., McHenry, J. J., Corpe, V. A., Rose, S. R., Lammlein, S. E., Kemery, E., Borman, W. C., Mendel, R., & Bosshardt, M. J. (1988). Development and field test of behaviorally anchored scales for nine MOS (ARI Technical Report 776). (AD A194 271)

Wallace, S. R. (1965). Criteria for what? American Psychologist, 20, 411-418.

Wing, H., Peterson, N. G., & Hoffman, R. G. (1984). Expert judgments of predictor-criterion validity relationships. Paper presented at the Annual Convention of the American Psychological Association, Toronto.

Wise, L. L., Campbell, J. P., McHenry, J. J., & Hanser, L. M. (1986). A latent structure model of job performance factors. Paper presented at the Annual Convention of the American Psychological Association, Washington.

Wise, L. L., McHenry, J. J., & Young, W. Y. (1986). Project A Concurrent Validation: Treatment of missing data (ARI Working Paper RS-WP-86-08).

Wise, L. L., & McLaughlin, D. H. (1980). Guidebook for the imputation of missing data. Palo Alto, California: American Institutes for Research.

Young, W. Y., Harris, J. H., Hoffman, G. R., & Houston, J. S. (In preparation). Large scale data collection and data base preparation. Paper prepared for the Annual Conference of the Society of Industrial and Organizational Psychology, Atlanta, April 1987.

APPENDIX A

PROJECT A TECHNICAL PAPERS FOR FISCAL YEAR 1986

APPENDIX A

PROJECT A TECHNICAL PAPERS FOR FISCAL YEAR 1986

A number of technical papers dealing with specialized aspects of Project A were prepared during Fiscal Year 1986. These papers are available in an ARI Research Note, Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1986 Fiscal Year - Supplement to ARI Technical Report ____ (in preparation). The following papers are included in the Research Note:

Arabian, J.M., & Hanser, L.M. (1986, August). Standard setting procedures: Army enlistment standards and job performance. Paper presented at the Annual Convention of the American Psychological Association, Washington.

Arabian, J., Rumsey, M., & McHenry, J. (1986). Army research to link standards for enlistment to on-the-job performance. Army Research Institute Working Paper.

Borman, W.C. (1986, August). Performance criterion measurement: What are the different methods measuring? Paper presented at the Air Force Conference on Job Performance Measurement, Air Force Human Resources Laboratory, San Antonio.

Borman, W.C., Pulakos, E.D., & Motowidlo, S.J. (1986, August). Toward a general model of soldier effectiveness. Paper presented at the Annual Convention of the American Psychological Association, Washington.

Campbell, J.P. (1986, August). Project A: When the textbook goes operational. Paper presented at the Annual Convention of the American Psychological Association, Washington.

Eaton, N. K., Wing, H., & Lau, A. (1985, October). Utility estimation in five enlisted occupations. Paper presented at the Annual Conference of the Military Testing Association, San Diego.

Ford, P., Campbell, C.H., Felker, D.B., & Edwards, D.C. (1986, August). Comparability of hands-on and knowledge tests across nine military jobs. Paper presented at the Annual Convention of the American Psychological Association, Washington.

Hanser, L.M., & Arabian, J.M. (1985, October). Multi-dimensional performance measurement. Paper presented at the Annual Conference of the Military Testing Association, San Diego.

Hough, L.M., Gast, I.F., & White, L.A. (1986, August). The relation of leadership and individual differences to job performance. Paper presented at the Annual Convention of the American Psychological Association, Washington.

Hough, L.M., McGue, M.K., Kamp, J.D., Houston, J.S., & Barge, B.N. (1985, October) Measuring personal attributes: Temperament, biodata, and interests. Paper presented at the Annual Conference of the Military Testing Association, San Diego.

Humphreys, L.G. (1986, August). Stability and instability of individual differences. Paper presented at the Annual Convention of the American Psychological Association, Washington.

McHenry, J.J., & Toquam, J.L. (1985, October). Computerized assessment of perceptual and psychomotor abilities. Paper presented at the Annual Conference of the Military Testing Association, San Diego.

Olson, D.M., & Borman, W.C. (1985, October). Examination of environmental determinants of Army performance criteria. Paper presented at the Annual Conference of the Military Testing Association, San Diego.

Olson, D.M., Borman, W.C., & Motowidlo, S.J. (1986, August). Individual differences and environmental determinants of Army performance criteria. Paper presented at the Annual Convention of the American Psychological Association, Washington.

Peterson, N.G. (1985, October). Mapping predictors to criterion space: Overview. Paper presented at the Annual Conference of the Military Testing Association, San Diego.

Rosse, R.L., & Peterson, N. (1985, October). Using microcomputers for assessment: Practical problems and solutions. Paper presented at the Annual Conference of the Military Testing Association, San Diego.

Rossmeissl, P.G., McLaughlin, D.H., Wise, L.L., & Brandt, D.A. (1985, October). The validity of ASVAB for predicting training and SQT performance. Paper presented at the Annual Conference of the Military Testing Association, San Diego.

Sadacca, R., de Vera, M.V., DiFazio, A., & White, L.A. (1986, August). Weighting performance constructs in composite measures of job performance. Paper presented at the Annual Convention of the American Psychological Association, Washington.

Smith, E.P. (1935, October). Developing new attribute requirements scales for military jobs. Paper presented at the Annual Conference of the Military Testing Association, San Diego.

Toquam, J.L., Dunnette, M.D., Corpe, V.A., & Houston, J. Adding to the ASVAB: Cognitive paper-and-pencil measures. Paper presented at the Annual Conference of the Military Testing Association, San Diego.

Walker, C.B. (1985, October). Three variables that may influence the validity of biodata. Paper presented at the Annual Conference of the Military Testing Association, San Diego.

White, L.A., Gast, I.F., & Rumsey, M.G. (1985, October). Leaders' behavior and the performance of first-term soldiers. Paper presented at the Annual Conference of the Military Testing Association, San Diego.

White, L.A., Borman, W.C., & Hough, L.M. (1986, August). A path analytic model of job performance ratings. Paper presented at the Annual Convention of the American Psychological Association, Washington.

Wing, H., Barge, B.N., & Hough, L.M. (1985, October). Vocational interests as predictors of Army performance. Paper presented at the Annual Conference of the Military Testing Association, San Diego.

Wise, L.L., Campbell, J.P., McHenry, J.J., & Hanser, L.M. (1986, August). A latent structure model of job performance factors. Paper presented at the Annual Convention of the American Psychological Association, Washington.

APPENDIX B

TASK CONTENTS OF JOB KNOWLEDGE TESTS, HANDS-ON TESTS, AND MOS-SPECIFIC TASK RATINGS FOR NINE MOS

APPENDIX B

TASK CONTENTS OF JOB KNOWLEDGE TESTS, HANDS-ON TESTS, AND MOS-SPECIFIC TASK RATINGS FOR NINE MOS

MOS 11B - INFANTRYMAN	Hands-On Test and Rating Scale	Knowledge Test
Put on and wear an M17-series protective mask	x	x
Perform CPR on an adult (one-man method)		x
Put on and wear protective clothing in accordance with established MOPP levels		x
Administer nerve agent antidote to self (self-aid)		x
Put on a field or pressure dressing	x	x
Collect/report information-SALUTE		x
Load, reduce stoppage, and clear an M60 MG	x	x
Perform operator maintenance on M16A1 rifle	x	x
Recognize and identify friendly and threat armored vehicles		x
Set headspace and timing on a caliber .50 MG[a]	x	x
Camouflage yourself and your individual equipment		x
Move under direct fire		x
Install and fire/recover an M18A1 Claymore mine	x	x
Establish an observation post		x
Select fire team/scout overwatch position		x
Select hasty firing positions in urban terrain		x
Techniques of movement in urban terrain	x	x
Estimate range		x
Conduct day and night surveillance w/o electronic devices	x	
Operate radio set AN/PRC-77 or AN/PRC-25 (AN/GRC-160 on AN/GRC-125 assembled for manpack operations)	x	x
Zero AN/PVS-4 on M16A1	x	x
Place an AN/PVS-5 (night vision goggles) into operation		x
Engage enemy targets with hand grenades	x	x
Prepare a Dragon for firing	x	x
Prepare a range card for an M60 machinegun	x	x
Engage targets with an M72A2 LAW	x	
Call for/adjust indirect fire		x
Navigate from one point on the ground to another point		x
Move over, through, or around obstacles (except minefields)		x
Identify terrain features on a map		x

[a]Not administered to Non-mechanized Infantry.

MOS 13B - CANNON CREWMAN	Hands-On Test and Rating Scale	Knowledge Test
Load, reduce a stoppage, and clear M16 rifle	x	x
Put on and wear an M17-series protective mask	x	x
Perform CPR on an adult (one-man method)	x	x
Decontaminate your skin		x
Put on and wear protective clothing in accordance with established MOPP levels	x	x
Administer nerve agent antidote to self (self-aid)	x	x
Camouflage equipment		x
Use challenge and password		x
Recognize and identify friendly and threat armored vehicles		x
Set headspace and timing on a caliber .50 MG	x	x
Prevent shock		x
Determine an azimuth using an M2 compass	x	x
Measure an azimuth on a map with a protractor	x	x
Use visual signals to control movement (mounted)	x	x
Perform operator maintenance on a caliber .50 HB machinegun and ammunition	x	x
Install and operate field telephone	x	x
Prepare DA Form 2404 (Equipment inspection and maintenance worksheet)		x
Perform PMCS[a]		x
Prepare a position to receive/emplace howitzer[b]		x
Drive SP howitzer or prime mover[b]		x
Prepare howitzer for operation[c]		x
Prepare ammunition for firing[d]		x
Store ammunition in preparation for firing		x
Emplace/recover collimator[c]	x	x
Emplace/recover aiming posts	x	x
Boresight the direct fire telescope using a distant aiming point (DAP)[a]	x	x
Sight on a target with direct fire telescope[a]	x	x
Load and fire a prepared round[a]		x
Lay howitzer for initial direction of fire	x	x
Disassemble/assemble breech and firing[a] mechanism	x	x

[a] Four tracked versions prepared for M109, M110, M198, and M102 howitzer crewmen.

[b] Two tracked versions prepared for self-propelled (M109, M110) and towed (M198, M102) howitzer crewmen.

[c] Three tracked versions prepared for M109/M110, M198, and M102 howitzer crewmen.

[d] Two tracked versions prepared for M109/M110/M198 and M102 howitzer crewmen.

[e] Two hands-on tracked versions prepared for M109/M110/M102 and M198 howitzer crewmen.

MOS 19E - ARMOR CREWMAN	Hands-On Test and Rating Scale	Knowledge Test
Put on and wear protective clothing in accordance with established MOPP levels		x
Administer nerve agent antidote to self (self-aid)		x
Determine grid coordinates of a point on a map using the military grid reference system	x	x
Put on a field or pressure dressing	x	x
Collect/report information-SALUTE		x
Recognize and identify friendly and threat armored vehicles		x
Prevent shock		x
Perform operator maintenance on a caliber .45 pistol		x
Know your rights and obligations as a prisoner of war		x
Install and fire/recover an M18A1 Claymore mine		x
Perform operator maintenance on M3/M3A1 submachinegun	x	x
Identify minefield markers		x
Identify terrain features on a map		x
Put on, wear, remove and store M24, M25, or M25A1 protective mask with hood	x	
Prepare a vehicle for nuclear attack		x
Operate gas particulate filter unit on M60-series tank	x	x
Escape from an M48A5/M60-series tank	x	x
Use an automated CEOI	x	x
Send a radio message	x	x
Operate radio set AN/VRC-64 or AN/GRC-160 (AN/VRC-53 or AN/GRC-125)	x	x
Engage targets with the main gun from the gunner's station on an M60A3 tank[a]		x
Perform misfire procedures on the main gun		x
Boresight and system calibrate an M60A3 tank[b]	x	x
Load/unload 105-mm main gun on an M48A5/M60-series tank		x
Perform operator maintenance on an M240 machinegun	x	x
Perform gunner's and loader's preventive maintenance prepare-to-fire checks and services on an M60A3 tank[a]	x	x
Prepare loader's station for operation on an M48A5/M60-series tank[b,c]	x	x
Start/stop the engine on an M48A5/M60-series tank	x	x
Extinguish a fire on an M48A5/M60-series tank		x
Remove/install track blocks on an M48A5/M60-series tank	x	x

[a]Not administered to M60A1 tank crewmen.

[b]Knowledge test not administered to M60A1 tank crewmen.

[c]Two tracked versions of hands-on test for M60A3 and M60A1 tank crewmen.

MOS 31C - SINGLE CHANNEL RADIO OPERATOR	Hands-On Test and Rating Scale	Knowledge Test
Handle classified equipment and material		x
Erect a doublet antenna		x
Troubleshoot radio teletypewriter set AN/GRC-142 or AN/GRC-122		x
Perform operator troubleshooting procedures on radio set AN/GRC-106		x
Erect, dismantle, and adjust expedient radio antennas		x
Perform operator's PMCS on terminal communications AN/UGC-74A(V)3		x
Operate in radio nets		x
Operate radio set control group AN/GRA-6		x
Maintain an M17-series protective mask		x
Know your rights and obligations as a prisoner of war		x
Decontaminate your skin and personal equipment		x
Practice noise, light, and litter discipline		x
Recognize and identify friendly and threat armored vehicles		x
Perform cardiopulmonary resuscitation (CPR) on an adult using one-man method		x
Perform operator maintenance on an M16A1 rifle, magazine, and ammunition		x
Load, reduce a stoppage and clear M16A1 rifle	x	x
Put on and wear protective clothing in accordance with established MOPP levels	x	x
Perform PMCS on M884 or M1028 trucks[a]	x	x
Establish, enter or leave a radio net	x	x
Operate generator set PU-620	x	x
Perform operator troubleshooting procedures on generator set PU-620	x	x
Use the KTC 1400D numerical cipher/ authentication system	x	x
Recognize electronic countermeasures and implement electronic counter-countermeasures	x	x
Prepare a message for transmission in 16-line format	x	x
Operate terminal communications AN/UGC-74A(V)3	x	x
Install radio teletypewriter set AN/GRC-142 or AN/GRC-122	x	x
Install radio set AN/GRC-106[a]	x	x
Operate radio teletypewriter set AN/GRC-142 or AN/GRC-122[b]	x	x
Determine the grid coordinates of a point on a military map using the military grid reference system	x	x
Put on a field or pressure dressing	x	x

[a] Three tracked versions for hands-on test.

[b] Two tracked versions for hands-on test.

MOS 63B - LIGHT WHEEL VEHICLE MECHANIC	Hands-On Test and Rating Scale	Knowledge Test
Perform expedient repairs		x
Slave start disabled vehicle		x
Troubleshoot engine cooling system (truck, cargo, 2 1/2-ton, 6x6)		x
Troubleshoot engines (truck, cargo, 2 1/2-ton, 6x6)		x
Troubleshoot steering system		x
Camouflage equipment		x
Administer first aid to a nerve agent casualty (buddy-aid)		x
Determine the grid coordinates of a point on a military map using military grid reference system		x
Perform operator maintenance on an M16A1 rifle, magazine, and ammunition		x
Put on and wear protective clothing in accordance with established mission-oriented protective posture (MOPP) levels		x
Replace radiator (truck, cargo, 2 1/2 ton, 6x6)		x
Replace starter (truck, cargo, 5-ton, 6x6)		x
Tow disabled vehicle with 5-ton wrecker		x
Perform annual PMCS (inspect and road test) (truck, cargo, 5-ton, 6x6)		x
Troubleshoot brake system malfunctions (truck, cargo, 1 1/4-ton, 4x4)		x
Use challenge and password	x	x
Determine a magnetic azimuth using a compass	x	x
Load, reduce stoppage, clear M16A1 rifle	x	x
Put on a field or pressure dressing	x	x
Put on, wear and remove your M17-series protective mask with hood	x	x
Adjust clutch pedal free travel (truck, utility, 1/4-ton, 4x4)	x	x
Maintain assigned toolkit[a]	x	x
Repair electrical wiring (truck, cargo, 1 1/4-ton, 4x4)	x	x
Replace air hydraulic cylinder (truck, cargo, 2 1/2-ton, 6x6)	x	x
Replace fuel pump (truck, cargo, 2 1/2-ton, 6x6)[a]	x	x
Replace wheel bearings (truck, cargo, 2 1/2-ton, 6x6)	x	x
Replace service brakes (truck, utility, 1/4-ton, 4x4)	x	x
Troubleshoot electrical system (truck, cargo, 5-ton, 6x6)[a]	x	x
Troubleshoot fuel system malfunctions (truck, cargo, 2 1/2-ton, 6x6)	x	x
Troubleshoot service brake malfunctions (truck, utility, 1/4-ton, 4x4)	x	x

[a] Two tracked versions for hands-on test.

MOS 64C - MOTOR TRANSPORT OPERATOR	Hands-On Test and Rating Scale	Knowledge Test
Load, reduce a stoppage, and clear M16 rifle	x	x
Put on and wear an M17-series protective mask	x	x
Perform CPR on an adult (one-man method)	x	x
Decontaminate your skin		x
Put on and wear protective clothing in accordance with established MOPP levels	x	x
Administer nerve agent antidote to self (self-aid)	x	x
Determine grid coordinates of a point on a map using the military grid reference system	x	x
Put on a field or pressure dressing	x	x
Camouflage equipment		x
Use challenge and password		x
Collect/report information-SALUTE		x
Drive vehicle in motor march or convoy		x
Administer first aid to nerve agent casualty (buddy aid)	x	x
Load, reduce stoppage, and clear an M60 MG	x	x
Perform PMCS (5 ton)		x
Perform operator maintenance on M16A1 rifle	x	x
Measure distance on a map	x	x
Visually identify threat aircraft		x
Decontaminate equipment using the ABC M11 decontaminating apparatus	x	x
Use M8 paper to identify a chemical agent	x	x
Drive vehicle off-road		x
Operate vehicle in snow and ice		x
Couple semitrailer[a]	x	
Uncouple semitrailer[a]	x	
Operate tractor and semitrailer[a]	x	x
Transport general cargo		x
Fill out SF 91 (Operator's Report of Motor Vehicle Accident)		x
Perform vehicle emergency recovery procedures		x
Drive under blackout		x
Use proper defense procedures when ambushed or attacked		x

[a]Two tracked versions for hands-on tests.

MOS 71L - ADMINISTRATIVE SPECIALIST	Hands-On Test and Rating Scale	Knowledge Test
Type a joint messageform	x	x
File documents and correspondence	x	x
Type a memorandum	x	x
Type a basic comment to a disposition form	x	x
Type a second or subsequent comment to a disposition form	x	x
Type a military letter	x	x
Receipt/transfer classified material	x	x
Receive, maintain, control office equipment		x
Type military orders	x	x
Establish functional files		x
Safeguard "For Official Use Only" material		x
Assemble correspondence	x	x
Camouflage yourself and individual equipment		x
Dispatch outgoing distribution		x
Control expendable/nonexpendable supplies		x
Know your rights and obligations as a prisoner of war		x
Practice noise, light, and litter discipline		x
Perform operator maintenance on an M16A1 rifle, magazine and ammunition	x	x
Determine the grid coordinates of a point on a military map using the military grid reference system	x	x
Put on, wear and remove your M17-series protective mask with hood	x	x
Maintain an M17-series protective mask		x
Administer nerve agent antidote to self (self-aid)		x
Load, reduce a stoppage and clear an M16A1 rifle		x
Determine a magnetic azimuth using a compass		x
Put on and wear protective clothing in accordance with established MOPP levels		x
Type straight copy	x	
Prepare requisition forms/publications	x	
Put on field or pressure dressing	x	

MOS 91A - MEDICAL SPECIALIST	Hands-On Test and Rating Scale	Knowledge Test
Draft/file TPR charts (SF 511)		x
Conduct PMCS on ambulance		x
Replace filters in an M17-series protective mask		x
Move over, through, or around obstacles (except minefields)		x
Evaluate a casualty		x
Practice noise, light and litter discipline		x
Prevent shock		x
Put on and wear protective clothing in accordance with established MOPP levels		x
Decontaminate your skin and personal equipment		x
Load, reduce stoppage, and clear M16A1 rifle		x
Recognize and identify friendly and threat armored vehicles		x
Triage: Establish priorities for treatment/ evacuation of casualties		x
Initiate treatment for shock (hypovolemic)		x
Assist with sick call procedures		x
Manage a patient with an intravenous infusion		x
Splint a suspected fracture	x	x
Determine the grid coordinates of a point on a military map using the military grid reference system	x	x
Perform cardiopulmonary resuscitation (CPR) on an adult using the one-man method	x	x
Put on a field or pressure dressing	x	x
Open the airway	x	x
Initiate a field medical card	x	x
Decontaminate mercury thermometers	x	x
Initiate an intravenous infusion[a]	x	x
Measure and record a patient's pulse	x	x
Measure and record a patient's respirations	x	x
Assemble a needle and syringe and draw medication[a]	x	x
Measure and record a patient's blood pressure	x	x
Change a sterile dressing	x	x
Administer an injection (intramuscular, intradermal)[a]	x	x
Establish and maintain a sterile field	x	x

[a]Two tracked versions for hands-on test.

MOS 95B - MILITARY POLICE	Hands-On Test and Rating Scale	Knowledge Test
Prepare MP reports and forms	x	x
Determine the grid coordinates of a point on a military map using the military grid reference system	x	x
Prepare/operate FM radio sets	x	x
Determine a magnetic azimuth using a compass	x	x
Use hand and arm signals to direct traffic	x	x
Operate a dismount point	x	x
Put on, wear and remove your M17-series protective mask with hood	x	x
Perform operator/crew PMCS on vehicle	x	x
Perform cardiopulmonary resuscitation (CPR) on an adult using the one-man method	x	x
Put on a field or pressure dressing	x	x
Load, reduce a stoppage, and clear an M16A1 rifle	x	x
Operate/maintain caliber .38 or .45 pistol[a]	x	x
Load, reduce a stoppage, and clear an M60 machinegun	x	x
Use automated CEOI		x
Load, reduce stoppage, and clear squad automatic weapon		x
Respond to a domestic disturbance		x
Move under direct fire		x
Perform patrol duties		x
Decide when to use force		x
Determine requirements for a lawful apprehension		x
Respond to and secure the scene of a traffic accident		x
React to hostile fire during convoy movement		x
Decontaminate your skin and your personal equipment		x
Maintain an M17-series protective mask		x
Camouflage yourself and your individual equipment		x
Recognize and identify friendly and threat armored vehicles		x
React to sniper fire		x
Collect and process evidence		x
Navigate from one position on the ground to another point	x	x
Call for and adjust indirect fire	x	x
Estimate range	x	x

[a]Two tracked versions.

APPENDIX C

STATISTICAL CHARACTERISTICS OF FUNCTIONAL CATEGORIES FOR HANDS-ON COMPONENTS AND JOB KNOWLEDGE COMPONENTS FOR NINE MOS

APPENDIX C

Statistical Characteristics of Functional Categories for Hands-On Components and Job Knowledge Components for Nine MOS

A. MOS 11B: Infantryman

					Correlation[b] With:							
Job Knowledge Component	Items	Mean	SD	Rel[a]	First Aid	NBC	Weapons	Navigate	Field Tech.	Comm.	ID Tgts.	Anti-Air/ Tank Wpns
First Aid	21	11.83	2.93	51	100							
NBC	11	7.79	2.42	71	37	100						
Weapons	40	25.31	5.85	78	51	52	100					
Navigate	24	12.02	3.77	68	44	51	59	100				
Field Techniques	70	40.94	8.60	82	48	58	66	62	100			
Communication	7	4.30	1.50	47	23	24	37	27	46	100		
Identify Targets	12	8.12	2.30	62	22	24	35	32	39	21	100	
Anti-Air/Tank Wpns	2	1.50	0.65	26	23	21	24	20	26	14	20	100

					Correlation[b] With:							
Hands-On Component	Steps	Mean	SD	Rel[a]	First Aid	NBC	Weapons	Navigate	Field Tech.	Comm.	ID Tgts.	Anti-Air/ Tank Wpns
First Aid	17	10.30	2.62	84	100			-			-	
NBC	8	12.16	2.20	73	15	100		-			-	
Weapons	138-146	45.85	8.72	92	20	10	100	-			-	
Field Techniques	42	38.62	5.86	60	14	05	29	-	100		-	
Communication	4	13.18	1.44	23	06	-03	-01	-	08	100	-	
Anti-Air/Tank Wpns	24	16.65	3.61	69	10	08	21	-	21	06	-	100

[a] Reliabilities were computed separately for tracks within categories; only the reliability for the track with the largest number of soldiers is reported here.

[b] Correlations are not adjusted for scale reliabilities. Decimals have been omitted from the correlations.

Statistical Characteristics of Functional Categories for Hands-On Components and Job Knowledge Components for Nine MOS

B. MOS 13B: Cannon Crewman

					Correlation[b] With:									
Job Knowledge Component	Items	Mean	SD	Rel[a]	First Aid	NBC	Weap	Nav	Field Tech	Customs & Laws	Comm	ID Tgts	Prepare Howitzer	Op. Howitzer Sights
First Aid	16	10.0	2.4	50	100									
NBC	17	13.3	3.0	73	42	100								
Weapons	20	12.4	3.3	64	33	41	100							
Navigate	7	4.2	1.8	66	29	37	32	100						
Field Techniques	13	9.5	1.8	54	16	34	27	28	100					
Customs & Laws	8	5.5	1.7	48	32	45	30	29	23	100				
Communications	2	1.1	0.7	12	09	15	12	12	09	13	100			
Identify Targets	12	7.0	2.3	53	15	23	13	28	17	16	09	100		
Prepare Howitzer	62-68	35.0	8.5	82	37	48	38	48	35	35	21	24	100	
Operate Howitzer Sights	19-23	12.7	3.2	67	26	43	34	34	33	28	18	18	58	100

					Correlation[b] With:									
Hands-On Component	Steps	Mean	SD	Rel[a]	First Aid	NBC	Weap	Nav	Field Tech	Customs & Laws	Comm	ID Tgts	Prepare Howitzer	Op. Howitzer Sights
First Aid	43	17.3	4.8	91	100					-		-		
NBC	44	22.7	3.6	85	04	100				-		-		
Weapons	78	27.0	10.3	97	10	09	100			-		-		
Navigate	11	14.0	5.8	70	03	11	06	100		-		-		
Field Techniques	10	7.5	2.6	64	-00	16	17	20	100	-		-		
Communications	14	10.6	1.6	62	13	06	08	04	13	-	100	-		
Prepare Howitzer	15-42	7.6	5.5	99	12	05	24	11	08	-	08	-	100	
Operate Howitzer Sights	79-83	40.2	12.9	94	10	14	34	25	19	-	09	-	39	100

a Reliabilities were computed separately for tracks within categories; only the reliability for the track with the largest number of soldiers is reported here.

b Correlations are not adjusted for scale reliabilities. Decimals have been omitted from the correlations.

Statistical Characteristics of Functional Categories for Hands-On Components and Job Knowledge Components for Nine MOS

C. MOS 19E: Armor Crewman

Job Knowledge Component	Items	Mean	SD	Rel[a]	First Aid	NBC	Weap	Nav	Field Tech	Customs & Laws	Comm	Identify Targets	Operate Tank	Tank Gunnery
					Correlation[b] With:									
First Aid	17	9.9	2.5	47	100									
NBC	14	11.0	2.7	75	33	100								
Weapons	22	15.7	2.9	55	38	46	100							
Navigate	14	9.5	2.3	64	41	45	41	100						
Field Techniques	19	11.5	2.7	58	39	41	43	40	100					
Customs & Laws	9	5.2	1.9	51	26	37	33	37	41	100				
Communications	21	11.1	3.5	67	28	44	41	39	37	33	100			
Identify Targets	12	10.0	1.8	64	26	26	18	29	21	20	16	100		
Operate Tanks	31	19.1	3.8	63	31	44	47	40	42	31	38	28	100	
Tank Gunnery	9-33	17.5	4.4	68	35	43	41	38	41	34	36	29	47	100

Hands-On Component	Steps	Mean	SD	Rel[a]	First Aid	NBC	Weap	Nav	Field Tech	Customs & Laws	Comm	Identify Targets	Operate Tank	Tank Gunnery
					Correlation[b] With:									
First Aid	17	11.5	1.9	77	100									
NBC	12	10.2	2.2	63	02	100								
Weapons	67	26.5	1.3	83	03	-04	100							
Navigate	12	11.4	2.2	75	26	10	09	100						
Communications	21	27.6	7.8	82	09	17	07	21	-	-	100			
Operate Tanks	27	41.8	5.9	64	05	16	21	07	-	-	21	-	100	
Tank Gunnery	25-80	33.5	6.3	84	14	18	10	17	-	-	18	-	31	100

a Reliabilities were computed separately for tracks within categories; only the reliability for the track with the largest number of soldiers is reported here.

b Correlations are not adjusted for scale reliabilities. Decimals have been omitted from the correlations.

Statistical Characteristics of Functional Categories for Hands-On Components and Job Knowledge Components for Nine MOS

D. MOS 31C: Single Channel Radio Operator

Job Knowl. Component	Items	Mean	SD	Rel	Correlation[a] With: First Aid	NBC	Weap	Nav	Field Tech	Customs & Laws	Comm	ID Tgts	Veh Op	Gen	TTY Sta Ops	Main TTY Eq.	Op TTY Eq.	Inst TTY Eq.
First Aid	20	10.7	2.7	41	100													
NBC	16	12.5	2.6	70	45	100												
Weapons	16	11.1	2.5	57	33	40	100											
Navigate	5	3.2	1.7	75	33	38	34	100										
Field Tech.	4	2.5	1.0	28	19	27	25	15	100									
Customs & Laws	8	5.4	1.5	50	30	39	25	41	27	100								
Communications	15	10.1	2.7	66	35	51	40	39	12	33	100							
Identify Tgts	12	6.7	2.1	49	21	17	16	21	11	16	16	100						
Vehicle Operation	9	4.5	1.8	44	21	27	18	17	12	11	28	15	100					
Generators	14	8.2	2.2	51	27	39	20	30	14	24	34	18	30	100				
TTY Station Ops.	23	12.8	3.3	61	36	47	35	36	26	33	49	23	35	42	100			
Maintain TTY Eq.	20	10.4	3.2	60	23	43	24	34	16	32	35	15	14	45	40	100		
Operate TTY Eq.	22	11.5	3.2	57	30	38	24	32	16	33	43	13	18	25	43	36	100	
Install TTY Eq.	22	13.8	3.6	70	37	50	43	39	35	37	49	29	31	45	54	46	44	100

Hands-On Component	Steps	Mean	SD	Rel	Correlation[a] With: First Aid	NBC	Weap	Nav	Field Tech	Customs & Laws	Comm	ID Tgts	Veh Op	Gen	TTY Sta Ops	Main TTY Eq.	Op TTY Eq.	Inst TTY Eq.
First Aid	17	9.9	2.3	76	100													
NBC	36	10.2	2.7	92	19	100												
Weapons	20	11.7	1.8	73	-05	13	100											
Navigate	12	9.3	3.3	85	22	11	07	100										
Communications	20	16.4	6.7	86	17	14	17	21	-	-	100							
Vehicle Operation	33	11.7	1.3	77	11	07	-01	09	-	-	10	-	100					
Generators	39	19.8	3.6	77	16	13	21	06	-	-	16	-	05	100				
TTY Station Ops.	33	16.2	3.7	76	08	12	16	22	-	-	31	-	04	20	100			
Operate TTY Eq.	112	19.8	4.6	96	-07	06	11	08	-	-	20	-	09	23	35	-	100	
Install TTY Eq.	31	22.2	2.4	60	10	09	10	24	-	-	13	-	11	26	23	-	25	100

[a] Correlations are not adjusted for scale reliabilities. Decimals have been omitted from the correlations.

Statistical Characteristics of Functional Categories for Hands-On Components and Job Knowledge Components for Nine MOS

E. MOS 63B: Light Wheel Vehicle Mechanic

Job Knowl. Component	Items	Mean	SD	Rel	Correlation[a] With: First Aid	NBC	Weap	Nav	Field Tech	Customs & Laws	Vehicle Op./ Recovery	Elec. System	Power Train/ Clutch	Fuel Cooling/ Lub.	Brake/ Steer/ Susp.
First Aid	16	10.6	2.0	34	100										
NBC	11	8.1	2.0	55	25	100									
Weapons	16	11.5	2.4	53	24	34	100								
Navigate	9	4.8	2.6	76	19	23	25	100							
Field Tech.	4	1.7	0.9	-02	04	17	16	11	100						
Customs & Laws	8	5.8	1.5	42	27	28	31	17	10	100					
Veh. Op./Recovery	28	15.6	4.2	68	25	41	39	30	18	30	100				
Electrical System	18	11.4	2.4	41	22	25	31	21	13	27	35	100			
Power Train/Clutch	14	10.3	2.3	63	31	38	38	26	13	35	52	43	100		
Fuel/Cooling/Lub.	29	20.0	4.8	78	23	41	56	25	24	35	53	39	50	100	
Brake/Steer/Susp.	41	25.5	5.7	77	25	44	53	27	21	32	62	43	55	69	100

Hands-On Component	Steps	Mean	SD	Rel	Correlation[a] With: First Aid	NBC	Weap	Nav	Field Tech	Customs & Laws	Vehicle Op./ Recovery	Elec. System	Power Train/ Clutch	Fuel Cooling/ Lub.	Brake/ Steer/ Susp.
First Aid	17	9.9	2.2	75	100										
NBC	8	12.0	2.1	68	08	100									
Weapons	20	12.4	1.2	67	-01	01	100								
Navigate	4	10.3	3.5	53	04	10	-01	100	-						
Customs & Laws	12	12.1	1.8	79	04	-01	05	01	-	100					
Veh. Op./Recovery	11	11.2	1.9	59	04	02	06	05	-	05	100				
Electrical System	21	25.7	2.4	51	04	01	06	07	-	-02	12	100			
Power Train/Clutch	10	11.6	2.2	69	-00	01	07	06	-	01	07	08	100		
Fuel/Cooling/Lub.	28	24.3	2.6	74	04	01	03	06	-	01	13	12	10	100	
Brake/Steer/Susp.	41	48.1	4.5	73	02	05	07	07	-	-02	07	13	33	21	100

a Correlations are not adjusted for scale reliabilities. Decimals have been omitted from the correlations.

Statistical Characteristics of Functional Categories for Hands-On Components and Job Knowledge Components for Nine MOS

F. MOS 64C: Motor Transport Operator

Job Knowledge Component	Items	Mean	SD	Rel	Correlation[a] With: First Aid	NBC	Weapons	Navigate	Field Tech.	Customs & Laws	ID Targets	Veh. Op.
First Aid	27	16.9	3.1	52	100							
NBC	23	16.2	3.6	72	41	100						
Weapons	18	11.0	2.5	46	40	44	100					
Navigate	7	3.4	1.8	60	33	41	35	100				
Field Techniques	11	6.9	2.4	64	36	51	32	33	100			
Customs & Laws	8	5.7	1.6	47	24	33	29	29	31	100		
Identify Targets	8	2.1	1.4	23	06	05	04	12	08	03	100	
Vehicle Operation	65	35.0	7.9	81	43	41	37	39	40	38	11	100

Hands-On Component	Steps	Mean	SD	Rel	Correlation[a] With: First Aid	NBC	Weapons	Navigate	Field Tech.	Customs & Laws	ID Targets	Veh. Op.
First Aid	79	37.6	6.8	90	100							
NBC	57	45.9	5.8	79	23	100						
Weapons	92	29.0	5.8	93	19	14	100					
Navigate	15	14.4	7.3	76	24	10	15	100				
Vehicle Operation	41	33.3	4.2	73	06	11	10	09	-	-	-	100

a Correlations are not adjusted for scale reliabilities. Decimals have been omitted from the correlations.

Statistical Characteristics of Functional Categories for Hands-On Components and Job Knowledge Components for Nine MOS

G. MOS 71L: Administrative Specialist

Job Knowl. Component	Items	Mean	SD	Rel	Correlation[a] With: First Aid	NBC	Weapons	Navigate	Field Tech	Customs & Laws	Forms/ Files Mng.	Sup./ Coord.	Corres.	Class. Material
First Aid	8	4.7	1.4	35	100									
NBC	16	11.3	2.4	54	20	100								
Weapons	16	11.0	2.2	40	29	32	100							
Navigate	9	4.6	2.4	72	16	23	25	100						
Field Tech.	10	4.4	1.7	42	25	30	27	17	100					
Customs & Laws	8	5.2	1.5	40	17	29	22	17	27	100				
Forms/Files Mng.	20	12.0	4.2	78	28	31	26	23	19	21	100			
Sup./Coordination	8	4.4	1.5	32	17	22	16	21	21	16	36	100		
Correspondence	37	20.8	4.8	68	26	34	31	30	24	25	58	40	100	
Classified Material	12	4.4	1.9	33	20	23	25	15	17	20	32	19	32	100

Hands-On Component	Steps	Mean	SD	Rel	Correlation[a] With: First Aid	NBC	Weapons	Navigate	Field Tech	Customs & Laws	Forms/ Files Mng.	Sup./ Coord.	Corres.	Class. Material
First Aid	17	8.7	3.1	84	100									
NBC	8	11.7	2.2	67	21	100								
Weapons	43	10.3	2.3	91	27	20	100							
Navigate	12	8.3	3.8	87	35	17	27	100						
Forms/Files Mng.	14	17.9	4.8	74	11	04	11	14	-	-	100			
Correspondence	80	63.3	10.8	79	13	03	11	20	-	-	39	-	100	
Classified Material	13	3.5	3.3	86	07	09	12	06	-	-	10	-	24	100

a Correlations are not adjusted for scale reliabilities. Decimals have been omitted from the correlations.

Statistical Characteristics of Functional Categories for Hands-On Components and Job Knowledge Components for Nine MOS

H. MOS 91A: Medical Specialist

Job Knowl. Component	Items	Mean	SD	Rel[a]	Correlation[b] With: First Aid	NBC	Weap	Nav	Field Tech	ID Targets	Vehicle Operation	Treatment/Care	Clinic Housekeep.	Clinic Mngmt.
First Aid	40	27.8	5.1	75	100									
NBC	20	14.5	3.3	70	55	100								
Weapons	7	5.0	1.5	48	41	43	100							
Navigate	5	3.5	1.5	71	33	30	24	100						
Field Tech.	13	6.6	2.0	38	29	31	26	21	100					
ID Targets	12	6.7	2.3	56	16	15	21	18	18	100				
Vehicle Operation	6	2.4	1.0	-09	12	11	10	06	04	12	100			
Treatment/Care	110	74.3	11.9	86	71	61	43	37	37	19	13	100		
Clinic Housekeep.	7	4.7	1.5	44	42	29	21	10	17	10	04	36	100	
Clinic Management	10	5.6	2.4	68	47	31	14	16	12	-02	-01	44	21	100

Hands-On Component	Steps	Mean	SD	Rel[a]	Correlation[b] With: First Aid	NBC	Weap	Nav	Field Tech	ID Targets	Vehicle Operation	Treatment/Care	Clinic Housekeep.	Clinic Mngmt.
First Aid	55	33.4	4.5	81	100									
Navigate	12	9.4	3.1	78	21	-	-	100						
Treatment/Care	138	100.3	11.2	81	32	-	-	23	-	-	-	100		
Clinic Housekeep.	15	7.2	3.6	85	-00	-	-	01	-	-	-	11	100	-

a Reliabilities were computed separately for tracks within categories; only the reliability for the track with the largest number of soldiers is reported here.

b Correlations are not adjusted for scale reliabilities. Decimals have been omitted from the correlations.

Statistical Characteristics of Functional Categories for Hands-On Components and Job Knowledge Components for Nine MOS

I. MOS 95B: Military Police

Job Knowl. Component	Items	Mean	SD	Rel	Correlation[a] With: First Aid	NBC	Weap	Nav	Field Tech	Comm.	ID Tgts	Veh. Op.	Respond to Alarm	Conduct MP Proc.	Patrol Duties
First Aid	20	12.7	2.7	50	100										
NBC	17	12.7	3.0	74	32	100									
Weapons	26-29	19.7	3.8	64	40	47	100								
Navigate	21	11.1	3.5	68	26	35	36	100							
Field Tech.	27	15.6	3.6	65	36	40	47	35	100						
Communications	23	13.5	4.6	81	24	40	34	45	39	100					
Identify Targets	12	6.9	2.3	58	17	19	17	14	29	19	100				
Vehicle Operation	4	2.0	1.2	37	18	17	25	19	27	21	15	100			
Respond to Alarms	16	10.7	2.6	58	36	38	48	26	48	32	17	20	100		
Conduct MP Proc.	11	7.9	1.6	44	25	21	30	14	33	20	11	06	38	100	
Patrol Duties	33	19.8	3.4	53	37	37	37	29	43	31	12	17	42	34	100

Hands-On Component	Steps	Mean	SD	Rel	Correlation[a] With: First Aid	NBC	Weap	Nav	Field Tech	Comm.	ID Tgts	Veh. Op.	Respond to Alarm	Conduct MP Proc.	Patrol Duties
First Aid	45	20.0	4.7	91	100										
NBC	8	11.6	2.1	57	07	100									
Weapons	88	35.2	3.3	84	09	11	100								
Navigate	27	35.2	7.1	58	13	09	18	100							
Field Techniques	20	2.7	2.0	76	07	06	12	17	100						
Communications	17	10.5	2.2	79	10	09	22	17	18	100					
Vehicle Operation	29	10.6	1.7	75	13	12	14	13	05	29	-	100			
Patrol Duties	69	31.5	4.6	84	11	05	07	16	09	09	-	08	-	-	100

a Correlations are not adjusted for scale reliabilities. Decimals have been omitted from the correlations.

APPENDIX D

JOB PERFORMANCE MEASURE SUMMARY STATISTICS

FOR NINE MOS

APPENDIX D

JOB PERFORMANCE MEASURE SUMMARY STATISTICS

FOR NINE MOS

TABLE 1

JOB PERFORMANCE MEASURE SUMMARY STATISTICS
FOR 11B: INFANTRY

#	VARIABLE	MN	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	Overall Rating	4.60	0.90	.	90	74	68	77	85	75	65	23	12	17	-35	36	26	14	4	35	25	11	10	33	19	18	12	14
2	Eff/Ldr Rating	4.41	0.82	90	.	74	65	80	88	80	67	24	8	13	-30	36	30	12	5	36	27	10	13	33	20	20	9	17
3	Discipline Rtng	4.50	0.87	74	74	.	49	55	71	63	66	13	3	7	-39	31	16	10	3	30	22	6	8	24	13	13	5	13
4	Fitness Rating	4.86	0.89	68	65	49	.	59	66	52	45	17	27	9	-24	22	10	9	-1	10	10	-2	-4	13	6	6	1	1
5	Job-Spec Tech	32.98	4.58	77	80	55	59	.	86	75	58	23	15	17	-20	22	27	15	5	35	22	12	10	36	21	23	9	16
6	Job-Spec Other	22.67	3.66	85	88	71	66	86	.	80	67	25	8	14	-28	32	23	10	6	35	26	12	12	33	17	22	11	17
7	Combat Exmplry	9.02	1.49	75	80	63	52	75	80	.	75	24	8	13	-31	29	28	12	7	37	25	9	16	34	22	23	9	19
8	Combat Problems	10.03	1.64	65	67	66	45	58	67	75	.	14	8	6	-33	27	20	7	-1	36	24	9	15	31	21	18	8	14
9	Awards & Certs	3.33	2.18	23	24	13	17	23	25	24	14	.	15	20	-2	4	13	6	-1	14	15	-0	13	9	9	5	4	12
10	Phys. Readiness	273.44	28.00	12	8	3	27	15	8	8	8	15	.	11	2	-6	1	-7	-9	0	5	-7	-0	8	-2	-1	-4	-6
11	M16 Qualific.	2.74	0.57	17	13	7	9	17	14	13	6	20	11	.	1	1	13	6	-0	10	2	3	0	14	10	5	3	6
12	Articles 15	0.39	0.85	-35	-30	-39	-24	-20	-28	-31	-33	-2	2	1	.	-45	-10	-1	-6	-10	-9	-6	-6	-10	-1	-9	0	-5
13	Promotion Rate	0.03	0.68	36	36	31	22	22	32	29	27	4	-6	1	-45	.	16	7	7	19	17	12	10	18	14	12	11	17
14	HO Basic	50.50	10.06	26	30	16	10	27	23	28	20	13	1	13	-10	16	.	15	6	44	30	13	27	40	24	20	16	30
15	HO Safety	22.67	3.41	14	12	10	9	15	10	12	7	6	-7	6	-1	7	15	.	2	16	8	1	8	16	7	3	3	4
16	HO Comm	13.15	1.53	4	5	3	-1	5	6	7	-1	-1	-9	-0	-6	7	6	2	.	4	6	-1	-3	0	4	6	2	-1
17	JK Basic	50.93	9.71	35	36	30	10	35	35	37	36	14	0	10	-10	19	44	16	4	.	68	40	42	65	50	40	30	35
18	JK Safety	20.02	4.31	25	27	22	10	22	26	25	24	15	5	2	-9	17	30	8	6	68	.	23	26	47	41	32	25	20
19	JK Comm	4.37	1.47	11	10	6	-2	12	12	9	9	-0	-7	3	-6	12	13	1	-1	40	23	.	16	26	25	19	14	18
20	JK Identify	8.25	2.24	10	13	8	-4	10	12	16	15	13	-0	0	-6	10	27	8	-3	42	26	16	.	31	24	18	16	37
21	SK Basic	72.87	14.89	33	33	24	13	36	33	34	31	9	8	14	-10	18	40	16	0	65	47	26	31	.	63	60	44	43
22	SK Safety	9.51	2.12	19	20	13	6	21	17	22	21	9	-2	10	-1	14	24	7	4	50	41	25	24	63	.	45	34	26
23	SK Comm	5.68	1.67	18	20	13	6	23	22	23	18	5	-1	5	-9	12	20	3	6	40	32	19	18	60	45	.	40	31
24	SK Vehicle	0.78	0.42	12	9	5	1	9	11	9	8	4	-4	3	0	11	16	3	2	30	25	14	16	44	34	40	.	21
25	SK Identify	2.80	1.16	14	17	13	1	16	17	19	14	12	-6	6	-5	17	30	4	-1	35	20	18	37	43	26	31	21	.

N= 503

TABLE 2

JOB PERFORMANCE MEASURE SUMMARY STATISTICS
FOR 13B: CANNON CREWMAN

#	VARIABLE	MN	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	Overall Rating	4.59	0.79	.	86	71	61	62	72	73	61	11	10	5	-25	30	20	19	17	6	26	18	14	8	3	24	15	12	8	9
2	Eff/Ldr Rating	4.43	0.76	86	.	75	62	65	74	78	61	14	6	1	-23	25	27	25	14	9	32	20	15	11	5	30	20	15	5	13
3	Discipline Rtng	4.61	0.78	71	75	.	51	53	60	63	60	-0	-4	-1	-20	26	12	9	12	4	22	16	15	4	3	18	14	14	6	16
4	Fitness Rating	4.95	0.82	61	62	51	.	47	53	51	39	7	23	-1	-25	16	8	4	0	3	5	-1	-1	1	-2	4	-4	-1	-4	-9
5	Job-Spec Tech	23.59	3.55	62	65	53	47	.	80	60	39	11	10	1	-2	10	35	18	9	-1	25	10	10	17	8	24	8	12	6	4
6	Job-Spec Other	23.90	3.08	72	74	60	53	80	.	66	49	6	5	-4	-9	18	25	18	8	1	29	18	15	13	6	26	14	16	4	8
7	Combat Exmplry	9.00	1.44	73	78	63	51	60	66	.	63	14	10	3	-15	23	20	23	13	3	22	16	13	6	8	23	12	7	-1	1
8	Combat Problems	9.92	1.56	61	61	60	39	39	49	63	.	8	7	-3	-16	26	14	16	8	12	19	17	10	14	8	15	14	9	5	8
9	Awards & Certs	2.58	1.82	11	14	-0	7	11	6	14	8	.	12	18	0	8	15	19	15	-1	11	10	6	5	8	11	6	5	8	2
10	Phys. Readiness	261.74	32.70	10	6	-4	23	10	5	10	7	12	.	11	-3	-2	7	2	-7	8	-8	-8	-10	5	4	-0	-8	-10	-12	-15
11	M16 Qualific.	2.25	0.69	5	1	-1	-1	1	-4	3	-3	18	11	.	6	1	7	8	12	-3	-4	-5	-6	7	-3	-3	-7	0	3	-3
12	Articles 15	0.46	1.03	-25	-23	-20	-25	-2	-9	-15	-16	0	-3	6	.	-31	-0	-4	-5	-5	-7	-10	-12	-7	1	-5	-6	-2	-5	-3
13	Promotion Rate	0.01	0.63	30	25	26	16	10	18	23	26	8	-2	1	-31	.	6	10	10	3	10	6	5	5	-1	2	5	-2	10	7
14	HO Tech.	50.71	9.94	20	27	12	8	35	25	20	14	15	7	7	-0	6	.	47	20	11	33	13	7	10	12	36	18	20	11	9
15	HO Basic	48.50	13.00	19	25	9	4	18	18	23	16	19	2	8	-4	10	47	.	21	8	42	38	20	9	15	40	25	17	15	9
16	HO Safety	40.16	6.23	17	14	12	0	9	8	13	8	15	-7	12	-5	10	20	21	.	11	24	14	11	9	3	25	20	18	11	24
17	HO Comm	10.60	1.59	6	9	4	3	-1	1	3	12	-1	8	-3	-5	3	11	8	11	.	1	1	-2	6	5	7	5	-1	1	3
18	JK Tech.	50.67	9.94	26	32	22	5	25	29	22	19	11	-8	-4	-7	10	33	42	24	1	.	58	54	21	20	64	52	41	37	35
19	JK Basic	31.91	5.78	18	20	16	-1	10	18	16	17	10	-8	-5	-10	6	13	38	14	1	58	.	55	14	23	52	49	38	35	27
20	JK Safety	23.58	4.43	14	15	15	-1	10	15	13	10	6	-10	-6	-12	5	7	20	11	-2	54	55	.	10	21	41	38	35	26	27
21	JK Comm	1.12	0.68	8	11	4	1	17	13	6	14	5	5	7	-7	5	10	9	9	6	21	14	10	.	13	19	13	16	14	11
22	JK Identify	7.12	2.25	3	5	3	-2	8	6	8	8	8	4	-3	1	-1	12	15	3	5	20	23	21	13	.	20	21	25	10	8
23	SK Tech.	50.82	9.84	24	30	18	4	24	26	23	15	11	-0	-3	-5	2	36	40	25	7	64	52	41	19	20	.	63	47	38	40
24	SK Basic	23.17	5.27	15	20	14	-4	8	14	12	14	6	-8	-7	-6	5	18	25	20	5	52	49	36	13	21	63	.	51	40	52
25	SK Safety	8.44	2.12	12	15	14	-1	12	16	7	9	5	-10	0	-2	-2	20	17	18	-1	41	38	35	16	25	47	51	.	28	36
26	SK Comm	3.55	1.21	8	5	6	-4	6	4	-1	5	8	-12	3	-5	10	11	15	11	1	37	35	26	14	10	38	40	28	.	32
27	SK Vehicle	2.75	1.07	9	13	16	-8	4	8	1	8	2	-15	-3	-3	7	9	9	24	3	35	27	27	11	8	40	52	36	32	.

N= 401

TABLE 3

JOB PERFORMANCE MEASURE SUMMARY STATISTICS
FOR 19E: ARMOR CREWMAN

#	VARIABLE	MN	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	Overall Rating	4.62	0.78	.	84	72	58	72	53	69	61	12	20	8	-37	41	12	15	16	4	25	23	22	19	10	27	26	18	22	18	7
2	Eff/Ldr Rating	4.38	0.74	84	.	68	55	76	50	80	65	16	21	8	-32	41	17	26	19	17	31	34	31	27	14	33	32	23	22	23	11
3	Discipline Rtng	4.50	0.83	72	68	.	45	53	41	55	64	-1	12	-14	-35	38	6	15	14	2	21	23	18	17	6	27	18	8	22	14	-2
4	Fitness Rating	4.76	0.82	58	55	45	.	44	39	43	36	10	43	-0	-19	28	8	2	16	-3	1	5	10	5	-4	0	-0	-0	4	2	2
5	Job-Spec Tech	23.19	3.20	72	76	53	44	.	75	71	55	10	14	17	-31	34	23	17	19	13	25	23	26	19	8	27	25	18	15	20	6
6	Job-Spec Other	14.71	1.89	53	50	41	39	75	.	50	41	9	13	13	-18	19	15	9	13	2	8	7	12	2	4	15	12	8	9	9	1
7	Combat Exmplry	8.98	1.36	69	80	55	43	71	50	.	63	15	18	8	-32	34	15	27	15	14	20	23	19	20	7	22	25	19	10	18	2
8	Combat Problems	9.80	1.47	61	65	64	36	55	41	63	.	-1	7	4	-31	29	13	22	13	6	24	18	21	13	8	24	18	17	15	16	-1
9	Awards & Certs	2.52	1.60	12	16	-1	10	10	9	15	-1	.	15	19	-7	13	6	4	-3	13	5	7	-0	10	-2	12	12	3	4	8	8
10	Phys. Readiness	249.41	27.11	20	21	12	43	14	13	18	7	15	.	-1	-10	10	-3	-3	4	2	-6	0	-6	4	1	-4	1	2	-2	2	-7
11	M16 Qualific.	2.40	0.68	8	8	-14	-0	17	13	8	4	19	-1	.	14	-1	7	7	3	10	11	12	13	17	31	10	6	12	2	16	-1
12	Articles 15	0.35	0.77	-37	-32	-35	-19	-31	-18	-32	-31	-7	-10	14	.	-43	-9	-8	-16	1	-13	-17	-17	-7	1	-19	-13	-13	-0	-7	-6
13	Promotion Rate	0.03	0.58	41	41	38	28	34	19	34	29	13	10	-1	-43	.	10	7	15	12	14	24	28	21	2	17	22	18	6	15	1
14	HO Tech.	50.00	9.99	12	17	6	8	23	15	15	13	6	-3	7	-9	10	.	18	24	20	36	27	27	13	18	23	18	9	2	19	0
15	HO Basic	38.16	2.48	15	26	15	2	17	9	27	22	4	-3	7	-8	7	18	.	21	23	30	32	25	21	18	21	25	11	4	19	-0
16	HO Safety	21.85	2.95	16	19	14	16	19	13	15	13	-3	4	3	-16	15	24	21	.	14	22	18	18	10	6	15	13	5	5	17	6
17	HO Comm	28.55	7.59	4	17	2	-3	13	2	14	6	13	2	10	1	12	20	23	14	.	23	28	25	32	11	20	23	13	3	23	3
18	JK Tech.	50.00	9.99	25	31	21	1	23	8	20	24	5	-6	11	-13	14	36	30	22	23	.	60	52	45	34	64	60	44	38	42	7
19	JK Basic	42.16	7.28	23	34	23	5	23	7	23	18	7	0	12	-17	24	27	32	18	28	60	.	65	53	30	65	67	46	41	43	6
20	JK Safety	21.19	4.10	22	31	18	10	26	12	19	21	-0	-6	13	-17	28	27	25	18	25	52	65	.	44	34	46	51	37	26	33	5
21	JK Comm	11.33	3.59	19	27	17	5	19	2	20	13	10	4	17	-7	21	13	21	10	32	45	53	44	.	16	45	51	34	30	24	2
22	JK Identify	10.05	1.78	10	14	6	-4	8	4	7	8	-2	1	31	1	2	18	18	6	11	34	30	34	16	.	24	28	22	18	37	3
23	SK Tech.	54.54	9.66	27	33	27	0	27	15	22	24	12	-4	10	-19	17	23	21	15	20	64	65	46	45	24	.	75	53	59	48	21
24	SK Basic	34.94	8.44	26	32	18	-0	25	12	25	18	12	1	6	-13	22	18	25	13	23	60	67	51	51	28	75	.	68	47	47	12
25	SK Safety	8.18	2.14	18	23	8	-0	18	8	19	17	3	2	12	-13	18	9	11	5	13	44	46	37	34	22	53	68	.	38	33	4
26	SK Comm	7.59	1.80	22	22	22	4	15	9	10	15	4	-2	2	-0	6	2	4	5	3	38	41	26	30	18	59	47	38	.	24	14
27	SK Vehicle	0.54	0.50	7	11	-2	2	6	1	2	-1	8	-7	-1	-6	1	0	-0	6	3	7	6	5	2	3	21	12	4	14	9	.
28	SK Identify	3.01	0.96	18	23	14	2	20	9	18	16	8	2	16	-7	15	19	19	17	23	42	43	33	24	37	48	47	33	24	.	9

N= 335

TABLE 4

JOB PERFORMANCE MEASURE SUMMARY STATISTICS
FOR 31C: SINGLE CHANNEL RADIO OPERATOR

#	VARIABLE	MN	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	Overall Rating	4.73	0.79	.	83	73	64	74	66	66	66	17	11	2	-31	30	20	24	15	15	-2	24	17	9	14	3	13	19	10	14	2	-2
2	Eff/Ldr Rating	4.48	0.72	83	.	68	57	81	71	68	63	18	12	7	-31	30	24	21	21	15	2	30	28	14	16	6	13	23	12	20	12	4
3	Discipline Rtng	4.64	0.88	73	68	.	52	54	58	53	60	4	4	-11	-32	26	10	14	7	10	-1	20	15	4	15	6	7	9	-4	10	-2	-8
4	Fitness Rating	5.05	0.88	64	57	52	.	47	40	42	42	11	34	-6	-25	24	12	8	10	4	-2	1	2	0	8	-4	6	-5	-6	-2	-9	-12
5	Job-Spec Tech	14.27	2.01	74	81	54	47	.	76	66	57	14	4	5	-16	22	20	24	20	11	-1	29	30	16	15	8	15	19	15	16	13	-1
6	Job-Spec Other	14.37	2.09	66	71	58	40	76	.	54	48	3	-3	-3	-17	22	11	18	15	1	0	17	22	8	9	2	5	5	2	9	3	-9
7	Combat Examplry	9.09	1.54	66	68	53	42	66	54	.	77	11	1	5	-21	17	6	13	18	23	-7	26	30	15	19	11	12	18	9	14	10	7
8	Combat Problems	10.47	1.71	66	63	60	42	57	48	77	.	9	-1	-2	-22	14	4	16	11	15	-0	22	24	3	14	-1	5	15	5	9	0	-3
9	Awards & Certs	2.16	1.75	17	18	4	11	14	3	11	9	.	23	10	2	12	9	12	6	3	2	10	10	11	-0	-5	8	8	12	4	4	6
10	Phys. Readiness	259.54	29.59	11	12	4	34	4	-3	1	-1	23	.	4	-11	4	1	-10	0	1	-6	-4	-8	4	1	3	-8	-4	-0	1	-13	-5
11	M16 Qualific.	2.16	0.77	2	7	-11	-6	5	-3	5	-2	10	4	.	4	3	4	5	10	7	5	7	10	8	-4	-6	5	9	4	4	11	10
12	Articles 15	0.34	0.84	-31	-31	-32	-25	-16	-17	-21	-22	2	-11	4	.	-34	-9	-3	-7	-12	-3	-16	-9	-13	-20	-10	-3	-11	-4	-12	-4	-3
13	Promotion Rate	-0.02	0.56	30	30	26	24	22	22	17	14	12	4	3	-34	.	8	12	21	9	5	18	17	10	19	13	12	13	15	12	4	-0
14	HO Tech.	78.44	9.49	20	24	10	12	20	11	6	4	9	1	4	-9	8	.	25	25	28	9	42	21	23	21	22	15	39	21	24	9	8
15	HO Basic	21.25	3.84	24	21	14	8	24	18	13	16	12	-10	5	-3	12	25	.	18	27	8	31	31	18	15	5	21	27	24	27	10	15
16	HO Safety	20.15	3.99	15	21	7	10	20	15	18	11	6	0	10	-7	21	25	18	.	23	16	10	21	13	9	6	8	11	10	19	4	9
17	HO Comm	16.73	6.59	15	15	10	4	11	1	23	15	3	1	7	-12	9	28	27	23	.	1	34	29	21	38	21	23	26	17	11	5	25
18	HO Vehicle	11.73	1.31	-2	2	-1	-2	-1	0	-7	-0	2	-6	5	-3	5	9	8	16	1	.	11	9	10	2	-6	7	22	16	14	12	11
19	JK Tech.	57.16	11.68	24	30	20	1	29	17	26	22	10	-4	7	-16	18	42	31	10	34	11	.	60	59	60	37	33	72	49	50	44	25
20	JK Basic	22.12	4.61	17	28	15	2	30	22	30	24	10	-8	10	-9	17	21	31	21	29	9	60	.	58	50	23	31	49	42	43	40	20
21	JK Safety	23.31	4.63	9	14	4	0	16	8	15	3	11	4	8	-13	10	23	18	13	21	10	59	58	.	50	28	30	44	40	48	36	24
22	JK Comm	10.12	2.74	14	16	15	8	15	9	19	14	-0	1	-4	-20	19	21	15	9	38	2	60	50	50	.	32	19	44	36	36	27	19
23	JK Vehicle	4.54	1.82	3	6	6	-4	8	2	11	-1	-5	3	-6	-10	13	22	5	6	21	-6	37	23	28	32	.	17	20	14	16	13	11
24	JK Identify	6.72	2.13	13	13	7	6	15	5	12	5	8	-6	5	-3	12	15	21	8	23	7	33	31	30	19	17	.	27	28	21	11	40
25	SK Tech.	77.87	15.43	19	23	9	-5	19	5	18	15	8	-4	9	-11	13	39	27	11	26	22	72	49	44	44	20	27	.	62	58	48	29
26	SK Basic	10.95	2.74	10	12	-4	-8	15	2	9	5	12	-0	4	-4	15	21	24	10	17	16	49	42	40	36	14	28	62	.	56	42	26
27	SK Safety	11.08	2.81	14	20	10	-2	16	9	14	9	4	1	4	-12	12	24	27	19	11	14	50	43	48	36	16	21	58	56	.	41	27
28	SK Vehicle	3.83	1.84	2	12	-2	-9	13	3	10	0	4	-13	11	-4	4	9	10	4	5	12	44	40	36	27	13	11	48	42	41	.	22
29	SK Identify	1.16	0.93	-2	4	-8	-12	-1	-9	7	-3	6	-5	10	-3	-0	8	15	9	25	11	25	20	24	19	11	40	29	26	27	22	.

N= 239

TABLE 5

JOB PERFORMANCE MEASURE SUMMARY STATISTICS
FOR 63B: LIGHT WEIGHT VEHICLE MECHANIC

#	VARIABLE	MN	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	Overall Rating	4.55	0.84	.	86	75	57	75	75	68	65	20	7	-4	-24	24	11	-1	5	10	20	15	22	11	21	22	21	15	19
2	Eff/Ldr Rating	4.31	0.83	86	.	75	50	84	78	69	66	21	1	-5	-23	23	19	-1	3	12	23	16	27	18	26	22	19	14	22
3	Discipline Rtng	4.54	0.88	75	75	.	51	63	65	59	66	15	2	-8	-27	26	10	-5	7	9	11	5	19	3	14	23	20	13	14
4	Fitness Rating	4.82	0.86	57	50	51	.	38	49	44	41	13	31	2	-20	20	-2	-2	8	7	-0	2	8	-0	-2	16	14	13	8
5	Job-Spec Tech	22.42	4.10	75	84	63	38	.	78	65	57	21	-1	-5	-16	16	23	1	3	13	28	21	26	19	37	21	18	6	28
6	Job-Spec Other	23.19	3.52	75	78	65	49	78	.	68	55	18	5	-8	-18	17	12	4	4	12	18	17	22	13	21	18	18	9	20
7	Combat Exmplry	8.87	1.61	68	69	59	44	65	68	.	69	14	4	-7	-16	17	13	0	9	9	16	11	23	8	20	18	14	8	13
8	Combat Problems	9.92	1.86	65	66	66	41	57	55	69	.	14	-0	-6	-20	27	10	-3	4	9	17	11	20	7	19	21	18	18	18
9	Awards & Certs	2.31	1.81	20	21	15	13	21	18	14	14	.	4	2	-11	7	11	-5	-0	7	7	2	12	11	13	14	10	8	18
10	Phys. Readiness	255.47	31.93	7	1	2	31	-1	5	4	-0	4	.	10	-10	15	1	8	3	-1	-7	-12	-2	-9	-10	1	0	-3	-4
11	M16 Qualific.	2.19	0.73	-4	-5	-8	2	-5	-8	-7	-6	2	10	.	1	-9	-2	5	-4	-0	-6	3	3	2	-2	-2	2	-0	4
12	Articles 15	0.37	0.85	-24	-23	-27	-20	-16	-18	-16	-20	-11	-10	1	.	-36	-3	-2	-2	-4	-7	-5	-6	-0	-6	-11	-7	-13	-8
13	Promotion Rate	0.04	0.52	24	23	26	20	16	17	17	27	7	15	-9	-36	.	-5	-4	-2	-1	13	9	4	8	13	16	15	8	13
14	HO Tech.	110.11	6.84	11	19	10	-2	23	12	13	10	11	1	-2	-3	-5	.	8	6	18	33	23	19	22	37	19	16	4	24
15	HO Basic	34.96	4.09	-1	-1	-5	-2	1	4	0	-3	-5	8	5	-2	-4	8	.	10	7	6	12	14	12	10	7	15	-1	14
16	HO Safety	21.92	3.25	5	3	7	8	3	4	9	4	-0	3	-4	-2	-2	6	10	.	2	2	5	18	1	1	2	7	-7	-0
17	HO Vehicle	11.22	1.84	10	12	9	7	13	12	9	9	7	-1	-0	-4	-1	18	7	2	.	15	6	4	11	17	6	6	2	13
18	JK Tech.	68.61	11.93	20	23	11	-0	28	18	16	17	7	-7	-6	-7	13	33	6	2	15	.	62	47	62	67	50	39	36	59
19	JK Basic	24.36	4.69	15	16	5	2	21	17	11	11	2	-12	3	-5	9	23	12	5	6	62	.	45	44	47	41	36	22	44
20	JK Safety	18.91	3.05	22	27	19	8	26	22	23	20	12	-2	3	-6	4	19	14	18	4	47	45	.	38	40	36	33	20	39
21	JK Vehicle	15.81	4.03	11	18	3	-0	19	13	8	7	11	-9	2	-0	8	22	12	1	11	62	44	38	.	56	37	31	24	49
22	SK Tech.	56.00	12.89	21	26	14	-2	37	21	20	19	13	-10	-2	-6	13	37	10	1	17	67	47	40	56	.	52	47	30	69
23	SK Basic	16.56	4.24	22	22	23	16	21	18	18	21	14	1	-2	-11	16	19	7	2	6	50	41	36	37	52	.	61	50	56
24	SK Safety	6.02	1.74	21	19	20	14	18	18	14	18	10	0	2	-7	15	16	15	7	6	39	36	33	31	47	61	.	39	50
25	SK Comm	0.90	0.30	15	14	13	13	6	9	8	18	8	-3	-0	-13	8	4	-1	-7	2	36	22	20	24	30	50	39	.	39
26	SK Vehicle	24.10	5.54	19	22	14	8	28	20	13	18	18	-4	4	-8	13	24	14	-0	13	59	44	39	49	69	56	50	39	.

N= 403

TABLE 6

JOB PERFORMANCE MEASURE SUMMARY STATISTICS
FOR 64C: MOTOR TRANSPORT OPERATOR

#	VARIABLE	MN	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	Overall Rating	4.52	0.78	.	86	78	63	72	59	68	58	13	11	4	-30	33	8	20	15	16	17	19	3	13	12	16	14
2	Eff/Ldr Rating	4.36	0.75	86	.	77	59	78	69	74	58	17	9	6	-25	31	16	20	18	23	22	26	7	21	17	15	21
3	Discipline Rtng	4.53	0.81	78	77	.	52	67	51	58	54	10	3	-2	-29	35	4	14	14	15	15	19	1	16	12	15	16
4	Fitness Rating	4.74	0.87	63	59	52	.	54	39	46	35	3	28	3	-20	21	-2	8	14	5	6	4	2	5	6	7	-2
5	Job-Spec Tech	29.61	3.76	72	78	67	54	.	78	65	52	13	6	7	-21	25	9	16	19	17	20	19	5	16	17	12	15
6	Job-Spec Other	17.79	2.52	59	69	51	39	78	.	63	41	18	4	13	-15	19	12	11	16	17	16	19	4	13	17	7	14
7	Combat Exmplry	8.80	1.45	68	74	58	46	65	63	.	65	12	6	11	-21	22	20	19	16	20	15	20	5	18	8	10	15
8	Combat Problems	9.50	1.63	58	58	54	35	52	41	65	.	8	-3	2	-24	26	12	15	10	16	17	22	7	15	14	20	19
9	Awards & Certs	3.12	2.08	13	17	10	3	13	18	12	8	.	6	11	5	12	8	4	5	-3	-2	1	6	3	4	-1	2
10	Phys. Readiness	248.48	37.70	11	9	3	28	6	4	6	-3	6	.	3	-6	-1	-1	3	2	-4	-4	-4	2	-2	-5	0	-8
11	M16 Qualific.	2.09	0.75	4	6	-2	3	7	13	11	2	11	3	.	4	-5	9	13	5	7	5	3	-0	-1	-3	1	-1
12	Articles 15	0.46	0.98	-30	-25	-29	-20	-21	-15	-21	-24	5	-6	4	.	-36	-1	-11	-11	-7	-13	-12	0	-5	-8	-12	-4
13	Promotion Rate	-0.01	0.57	33	31	35	21	25	19	22	26	12	-1	-5	-36	.	10	9	10	9	12	11	5	11	9	8	11
14	HO Basic	43.44	10.16	8	16	4	-2	9	12	20	12	8	-1	9	-1	10	.	29	10	44	31	30	7	23	21	6	32
15	HO Safety	83.73	9.84	20	20	14	8	16	11	19	15	4	3	13	-11	9	29	.	14	27	31	24	4	24	19	14	24
16	HO Vehicle	33.30	4.19	15	18	14	14	19	16	16	10	5	2	5	-11	10	10	14	.	5	8	15	3	10	11	1	11
17	JK Basic	27.38	5.82	16	23	15	5	17	17	20	16	-3	-4	7	-7	9	44	27	5	.	67	54	10	47	29	20	49
18	JK Safety	33.42	5.42	17	22	15	6	20	16	15	17	-2	-4	5	-13	12	31	31	8	67	.	49	4	42	47	23	49
19	JK Vehicle	35.40	7.70	19	26	19	4	19	19	20	22	1	-4	3	-12	11	30	24	15	54	49	.	11	49	40	27	55
20	JK Identify	2.15	1.41	3	7	1	2	5	4	5	7	6	2	-0	0	5	7	4	3	10	4	11	.	17	10	-2	13
21	SK Basic	16.41	4.36	13	21	16	5	16	13	18	15	3	-2	-1	-5	11	28	24	10	47	42	49	17	.	56	43	68
22	SK Safety	6.44	1.93	12	17	12	6	17	17	8	14	4	-5	-3	-8	9	21	19	11	39	47	40	10	56	.	36	59
23	SK Comm	0.89	0.32	16	15	15	7	12	7	10	20	-1	0	1	-12	8	6	14	1	20	23	27	-2	43	36	.	37
24	SK Vehicle	55.72	10.07	14	21	16	-2	15	14	15	19	2	-8	-1	-4	11	32	24	11	49	49	55	13	68	59	37	.

N= 477

TABLE 7

JOB PERFORMANCE MEASURE SUMMARY STATISTICS
FOR 71L: ADMINISTRATIVE SPECIALIST

#	VARIABLE	MN	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	Overall Rating	4.92	0.85	.	83	71	57	72	63	63	59	20	24	4	-23	20	17	14	3	22	15	17	21	13	11	5	10
2	Eff/Ldr Rating	4.64	0.78	83	.	73	56	73	65	70	60	21	19	2	-19	19	25	14	2	29	17	18	28	17	9	7	11
3	Discipline Rtng	5.01	0.88	71	73	.	47	63	55	58	58	13	13	4	-27	19	20	10	-3	22	15	11	20	7	8	1	4
4	Fitness Rating	5.23	0.89	57	56	47	.	40	39	55	49	20	35	5	-23	20	3	7	-3	1	2	2	-1	0	-5	0	-2
5	Job-Spec Tech	19.88	2.73	72	73	63	40	.	76	54	50	8	7	-5	-21	21	24	8	-2	28	16	16	28	10	9	6	7
6	Job-Spec Other	18.57	3.13	63	65	55	39	76	.	50	46	10	13	-1	-21	17	22	13	1	22	15	16	26	8	9	10	8
7	Combat Exmplry	8.74	1.83	63	70	58	55	54	50	.	72	24	19	8	-15	18	9	20	11	13	23	17	14	13	8	8	23
8	Combat Problems	10.72	1.95	59	60	58	49	50	46	72	.	21	16	7	-22	13	12	14	6	11	26	12	15	13	9	1	14
9	Awards & Certs	2.62	1.73	20	21	13	20	8	10	24	21	.	17	20	-4	9	-0	10	-1	-0	5	11	-0	-2	-2	5	1
10	Phys. Readiness	260.40	33.39	24	19	13	35	7	13	19	16	17	.	11	-9	5	1	6	5	0	-5	8	5	4	12	2	8
11	M16 Qualific.	1.86	0.80	4	2	4	5	-5	-1	8	7	20	11	.	3	2	-4	12	8	-6	7	3	-3	2	-7	-1	13
12	Articles 15	0.22	0.62	-23	-19	-27	-23	-21	-21	-15	-22	-4	-9	3	.	-42	-13	-5	1	-10	-7	2	-10	-5	-5	-5	4
13	Promotion Rate	0.01	0.46	20	19	19	20	21	17	18	13	9	5	2	-42	.	12	5	2	6	6	9	5	7	6	4	-0
14	HO Tech.	86.09	14.26	17	25	20	3	24	22	9	12	-0	1	-4	-13	12	.	28	13	58	34	33	58	25	23	7	11
15	HO Basic	18.56	5.00	14	14	10	7	8	13	20	14	10	6	12	-5	5	28	.	43	29	48	35	23	26	17	6	23
16	HO Safety	20.54	4.00	3	2	-3	-3	-2	1	11	6	-1	5	8	1	2	13	43	.	11	28	23	7	13	10	0	17
17	JK Tech.	42.21	9.53	22	29	22	1	28	22	13	11	-0	0	-6	-10	6	58	29	11	.	47	48	73	42	24	17	17
18	JK Basic	25.23	5.16	15	17	15	2	16	15	23	26	5	-5	7	-7	6	34	48	28	47	.	50	40	44	27	27	28
19	JK Safety	16.24	3.01	17	18	11	2	16	16	17	12	11	8	3	2	9	33	35	23	48	50	.	43	38	32	19	25
20	SK Tech.	44.99	9.78	21	28	20	-1	28	26	14	15	-0	5	-3	-10	5	58	23	7	73	40	43	.	44	33	15	16
21	SK Basic	9.90	2.28	13	17	7	0	10	8	13	13	-2	4	2	-5	7	25	26	13	42	44	38	44	.	32	18	31
22	SK Safety	4.26	1.29	11	9	8	-5	9	9	8	9	-2	12	-7	-5	6	23	17	10	24	27	32	33	32	.	4	15
23	SK Comm	0.38	0.48	5	7	1	0	6	10	8	1	5	2	-1	-5	4	7	6	0	17	27	19	15	18	4	.	11
24	SK Vehicle	2.71	1.21	10	11	4	-2	7	8	23	14	1	8	13	4	-0	11	23	17	17	28	25	16	31	15	11	.

N= 353

TABLE 8

JOB PERFORMANCE MEASURE SUMMARY STATISTICS
FOR 91A: MEDICAL SPECIALIST

#	VARIABLE	MN	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	Overall Rating	4.61	0.82	.	86	78	60	67	62	71	70	22	15	-2	-29	32	17	6	13	28	24	25	4	8	28	12	15	6
2	Eff/Ldr Rating	4.40	0.77	86	.	76	56	73	67	73	71	24	13	-4	-30	33	20	9	19	26	25	21	-2	13	33	14	16	9
3	Discipline Rtng	4.54	0.91	78	76	.	47	60	47	58	69	12	7	-8	-29	31	15	11	13	28	21	20	-4	6	33	14	11	10
4	Fitness Rating	4.74	0.92	60	56	47	.	41	38	49	47	10	39	0	-20	18	3	-0	-0	3	7	4	7	1	-1	2	-4	-15
5	Job-Spec Tech	23.09	3.24	67	73	60	41	.	67	55	54	15	6	-1	-27	26	18	2	13	22	16	14	-3	3	32	5	15	7
6	Job-Spec Other	18.47	2.55	62	67	47	38	67	.	64	51	28	7	9	-17	27	10	6	16	18	25	20	5	15	23	11	18	16
7	Combat Exmplry	9.20	1.48	71	73	58	49	55	64	.	79	30	9	9	-20	26	16	10	15	22	25	22	1	18	28	20	17	12
8	Combat Problems	10.11	1.77	70	71	69	47	54	51	79	.	23	5	-5	-28	30	14	6	11	24	22	23	-1	9	32	28	16	12
9	Awards & Certs	3.04	2.01	22	24	12	10	15	28	30	23	.	14	34	-6	13	3	7	22	4	10	8	11	18	4	11	12	6
10	Phys. Readiness	255.71	31.94	15	13	7	39	6	7	9	5	14	.	17	-11	-2	4	-6	-5	-3	-7	-2	3	-5	-8	-3	-6	-7
11	M16 Qualific.	2.08	0.78	-2	-4	-8	0	-1	9	9	-5	34	17	.	-1	-4	3	0	8	-8	5	-7	-0	12	-4	-2	2	2
12	Articles 15	0.41	0.89	-29	-30	-29	-20	-27	-17	-20	-28	-9	-11	-1	.	-33	-10	1	-7	-10	-7	-6	12	-5	-13	-16	-6	-1
13	Promotion Rate	-0.00	0.58	32	33	31	18	26	27	26	30	13	-2	-4	-33	.	10	9	7	16	20	9	-9	11	18	11	14	11
14	HO Tech.	50.48	10.02	17	20	15	3	18	10	16	14	3	4	3	-10	10	.	16	34	39	27	30	2	13	44	8	23	14
15	HO Basic	9.57	3.00	6	9	11	-0	2	6	10	6	7	-6	0	1	9	16	.	17	21	37	21	9	14	17	18	22	11
16	HO Safety	33.52	4.30	13	19	13	-0	13	16	15	11	22	-5	8	-7	7	34	17	.	32	32	33	3	17	30	10	33	18
17	JK Tech.	85.32	13.71	28	26	28	3	22	18	22	24	4	-3	-8	-10	16	39	21	32	.	54	78	13	16	67	20	48	22
18	JK Basic	15.19	3.63	24	25	21	7	16	25	25	22	10	-7	5	-7	20	27	37	32	54	.	55	8	24	41	23	33	22
19	JK Safety	42.71	7.35	25	21	20	4	14	20	22	23	8	-2	-7	-6	9	30	21	33	78	55	.	12	16	55	21	49	21
20	JK Vehicle	2.42	1.04	4	-2	-4	7	-3	5	1	-1	11	3	-0	12	-9	2	9	3	13	8	12	.	10	2	-3	6	6
21	JK Identify	6.62	2.32	8	13	6	1	3	15	18	9	18	-5	12	-5	11	13	14	17	16	24	16	10	.	15	15	13	13
22	SK Tech.	91.65	17.57	28	33	33	-1	32	23	29	32	4	-8	-4	-13	18	44	17	30	67	41	55	2	15	.	24	52	36
23	SK Basic	2.04	0.78	12	14	14	2	5	11	20	28	11	-3	-2	-16	11	8	18	10	20	23	21	-3	15	24	.	26	14
24	SK Safety	5.77	1.56	15	16	11	-4	15	18	17	16	12	-8	2	-6	14	28	22	33	48	38	49	6	13	52	26	.	27
25	SK Vehicle	4.51	1.62	6	9	10	-15	7	16	12	12	6	-7	2	-1	11	14	11	18	22	22	21	6	13	36	14	27	.

N= 372

TABLE 9

JOB PERFORMANCE MEASURE SUMMARY STATISTICS
FOR 95B: MILITARY POLICE

#	VARIABLE	MN	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	Overall Rating	4.74	0.80	.	87	69	70	78	68	74	70	18	22	13	-28	21	15	18	8	4	1	12	10	8	4	9	8	19	8	7	6	8	-6
2	Eff/Ldr Rating	4.50	0.73	87	.	71	61	77	72	78	68	20	17	11	-22	19	14	21	10	10	1	10	13	7	7	15	10	18	13	12	8	9	-5
3	Discipline Rtng	4.71	0.77	69	71	.	46	65	48	55	63	6	7	4	-27	26	6	9	5	7	-3	11	10	6	8	12	12	15	18	14	6	10	-3
4	Fitness Rating	4.90	0.84	70	61	46	.	58	56	55	52	16	43	13	-26	16	9	12	7	5	2	0	3	-2	-0	3	-3	3	5	2	-1	1	-7
5	Job-Spec Tech	29.00	3.66	78	77	65	58	.	73	68	63	15	15	11	-19	17	12	19	8	12	2	14	16	9	5	9	7	17	11	12	5	5	-2
6	Job-Spec Other	23.60	3.10	68	72	48	56	73	.	71	61	32	18	27	-16	8	11	22	14	19	7	2	13	6	4	16	7	9	12	6	5	4	-2
7	Combat Exmpiry	9.56	1.36	74	78	55	55	68	71	.	79	19	19	16	-17	14	17	19	14	14	6	10	17	11	7	15	6	19	15	9	9	8	1
8	Combat Problems	10.45	1.53	70	68	63	52	63	61	79	.	15	15	15	-28	21	14	16	11	10	-0	16	18	17	11	14	10	19	15	9	6	13	0
9	Awards & Certs	3.17	2.09	18	20	6	16	15	32	19	15	.	20	26	-3	11	8	16	6	8	11	-11	2	-1	7	-0	9	4	4	10	4	9	-0
10	Phys. Readiness	251.75	32.78	22	17	7	43	15	18	19	15	20	.	13	-12	7	-1	6	4	2	4	-6	1	-3	-3	-2	-12	-2	-8	-4	-3	2	-5
11	M16 Qualific.	2.28	0.76	13	11	4	13	11	27	16	15	26	13	.	1	-3	4	6	5	4	6	-3	7	3	2	-9	-1	1	-0	-2	-2	-1	4
12	Articles 15	0.27	0.70	-28	-22	-27	-26	-19	-16	-17	-28	-3	-12	1	.	-39	-4	-0	-8	-3	2	-8	-6	-5	-2	0	0	-7	-6	-3	5	-7	6
13	Promotion Rate	0.01	0.47	21	19	26	16	17	8	14	21	11	7	-3	-39	.	4	4	6	1	-3	15	15	16	10	6	2	0	10	7	1	2	-1
14	HO Tech.	31.58	4.63	15	14	6	9	12	11	17	14	8	-1	4	-4	4	.	18	12	6	11	13	11	10	7	3	5	14	7	3	10	6	-0
15	HO Basic	50.04	10.28	18	21	9	12	19	22	19	16	16	6	6	-0	4	18	.	20	21	18	18	34	26	21	17	5	12	27	23	12	15	-7
16	HO Safety	31.76	5.16	8	10	5	7	8	14	14	11	6	4	5	-8	6	12	20	.	9	15	10	20	21	21	12	9	15	17	18	9	10	-6
17	HO Comm	10.57	2.17	4	10	7	5	12	19	14	10	8	2	4	-3	1	6	21	9	.	31	14	21	13	30	14	7	9	21	16	17	11	-10
18	HO Vehicle	10.56	1.63	1	1	-3	2	2	7	6	-0	11	4	6	2	-3	11	18	15	31	.	1	4	8	19	16	2	11	12	9	13	10	-1
19	JK Tech.	38.44	5.90	12	10	11	0	14	2	10	16	-11	-6	-3	-8	15	13	18	10	14	1	.	60	53	35	18	15	40	33	28	24	19	1
20	JK Basic	50.11	9.99	10	13	10	3	16	13	17	18	2	1	7	-8	15	11	34	20	21	4	60	.	60	51	32	22	38	49	46	35	31	4
21	JK Safety	25.52	4.55	8	7	6	-2	9	6	11	17	-1	-3	3	-5	16	10	26	21	13	8	53	60	.	40	24	20	36	37	33	27	28	0
22	JK Comm	13.54	4.62	4	7	8	-0	5	4	7	11	7	-3	2	-2	10	7	21	21	30	19	35	51	40	.	26	18	22	33	31	36	24	-3
23	JK Vehicle	2.03	1.19	9	15	12	3	9	16	15	14	-0	-2	-9	0	6	3	17	12	14	16	18	32	24	26	.	15	18	23	23	20	17	-4
24	JK Identify	6.88	2.29	8	10	12	-3	7	7	6	10	9	-12	-1	0	2	5	5	9	7	2	15	22	20	18	15	.	21	20	21	17	12	-2
25	SK Tech.	40.20	7.04	19	18	15	3	17	9	19	19	4	-2	1	-7	0	14	12	15	9	11	40	38	36	22	18	21	.	49	49	38	37	2
26	SK Basic	17.85	3.86	8	13	18	5	11	12	15	15	4	-8	-0	-6	10	7	27	17	21	12	33	49	37	33	23	20	49	.	60	40	44	-1
27	SK Safety	14.45	3.35	7	12	14	2	12	6	9	9	10	-4	-2	-3	7	3	23	16	16	9	28	46	38	31	23	21	49	60	.	39	40	-6
28	SK Comm	3.12	1.23	6	8	6	-1	5	5	9	6	4	-3	-2	5	1	10	12	9	17	13	24	35	27	36	20	17	38	40	39	.	32	-0
29	SK Vehicle	6.02	1.90	8	9	10	1	5	4	8	13	9	2	-1	-7	2	6	15	10	11	10	19	31	28	24	17	12	37	44	40	32	.	-1
30	SK Identify	0.29	0.51	-6	-5	-3	-7	-2	-2	1	0	-0	-5	4	6	-1	-0	-7	-6	-10	-1	1	4	0	-3	-4	-2	2	-1	-6	-0	-1	.

N= 506

END
DATE
FILMED
DTIC
10-88